ePact
Strengthening evaluation effectiveness and impact

UKaid
from the British people

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

Mobile phones, nutrition, and health in Tanzania: Cost-effectiveness endline report

Simon Batchelor, Nigel Scott, Anna Clements, and Tom Jones (Gamos)

Final version for publication 27/03/2020

Oxford Policy Management itad IDS Institute of Development Studies Gamos INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE IFPRI

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

Acknowledgements

The authors would like to thank all of the stakeholders who agreed to take part in this research and to be interviewed. We are grateful to the GSM Association (GSMA), Cardno, the mHealth Tanzania Public–Private Partnership (mHealth Tanzania-PPP), the Tanzania Food and Nutrition Centre (TFNC), the Tanzanian Ministry of Health, Community Development, Gender, Elderly and Children (MoHCDGEC), and the Global Alliance for Improved Nutrition (GAIN) teams for their ongoing cooperation, and to the UK Foreign, Commonwealth and Development Office (FCDO) and Oxford Policy Management (OPM) for their continued support and contributions to this project. Finally, we would like to thank all internal and external reviewers of draft versions of this report.

We are also extremely grateful for the ongoing collaborative partnership with the Institute of Development Studies (IDS), which is overseeing the evaluation's qualitative component, led by Inka Barnett, and the International Food Policy Research Institute (IFPRI) team overseeing the quantitative randomised control trial component, led by Daniel O. Gilligan. Both teams have provided invaluable feedback and support towards the design and tools, as well as comments on drafts of this report.

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 Simon Batchelor, Nigel Scott, Anna Clements, and Tom Jones (Gamos). For further information contact j.gordon@ids.ac.uk.

The contact point for the client is Luisa Odell [luisa.odell@fcdo.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

Disclaimer

This report has been prepared by the e-Pact consortium for the named client, for services specified in the Terms of Reference and contract of engagement. 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 UK Government; however, the views expressed do not necessarily reflect the UK Government's official policies.

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 Simon Batchelor, Nigel Scott, Anna Clements, and Tom Jones (Gamos). For further information contact j.gordon@ids.ac.uk.

The contact point for the client is Luisa Odell [luisa.odell@fcdo.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

Data management

All intellectual property rights in any materials produced from the evaluation (including publication of research findings and any other reports and data) remain the property of IDS and associated subcontracted collaborators. IDS and all subcontracted 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. FCDO has unlimited access to any material produced from the evaluation. In order to promote the use and uptake of the evaluation findings, and in line with FCDO'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. Following approval of the report by FCDO, the endline report will be made available on IFPRI's and IDS's websites.

Executive summary

mNutrition in Tanzania

The mNutrition initiative is a five-year global programme supported by the UK Foreign, Commonwealth and Development Office (FCDO), managed and supported by the GSM Association (GSMA), and implemented by in-country mobile network operators (MNOs) and third-party providers that seeks to use mobile technology to improve the health and nutritional status of children and adults in low-income countries around the world. The mNutrition initiative is implemented through existing mAgri and mHealth programmes in 12 countries throughout sub-Saharan Africa and South Asia. FCDO has committed to conducting an independent evaluation of the mNutrition initiative. Given the scale of the mNutrition service the decision was made to select two countries for inclusion in the evaluation: the mHealth service in Tanzania and the mAgri service in Ghana. The mNutrition intervention that is the focus of the evaluation in Tanzania, and of this report, is an integrated service that combines an existing SMS-based health communication campaign targeting pregnant women and mothers of young children, known as 'Healthy Pregnancy, Healthy Baby' (HPHB), with approximately 120 nutrition-focused SMS messages in Swahili. The combined service is called Wazazi Nipendeni and is managed by Cardno and delivered through a platform implemented by Rasello. The combined service sends out SMS messages timed to the stage of pregnancy or age of the child. The service is designed to provide relevant information for pregnant women and children up to the age of five in order to affect beliefs and behaviours in key nutrition-related areas, including infant and young child feeding (IYCF) and women's dietary diversity, in recipient households.

Evaluation design

The aim of the impact evaluation is to assess the impact, cost-effectiveness, and commercial viability of the mNutrition service. The evaluation is being conducted by a consortium of researchers from Gamos, the Institute for Development Studies (IDS), and the International Food Policy Research Institute (IFPRI), and it relies on a variety of different tools and methods to collect evidence on the impact of the mNutrition intervention in Tanzania. Broadly, the evaluation can be classified into three distinct but closely integrated components: a qualitative component, a quantitative component, and a business model and cost-effectiveness component. This cost-effectiveness report draws on the quantitative component, which employed a cluster randomised controlled trial to identify the causal effect of the service on nutrition knowledge, IYCF practices, women's dietary diversity, and the nutritional status of young children. Surveyed households in villages randomly assigned to the treatment group were offered access to the mNutrition content on a mobile phone, free of charge, through a door-to-door, in-person visit; households in villages randomly assigned to the control group did not receive any offer of access to the service. Baseline and endline surveys were conducted at a two-year interval.

In practice, the quantitative component (Gilligan *et al*., forthcoming, 2020) estimates the impact of the mNutrition service on the primary and secondary outcomes in two ways. The main estimates are based on the comparison of outcomes across households in villages that were randomly assigned to receive the offer of access to the mNutrition content—the treatment group—and households in villages that were randomly assigned not to receive that offer—the control group. These intent-to-treat (ITT) estimates measure the impact of the offer of access to the mNutrition service on outcomes. ITT estimates are the preferred measures of service impact as they rely only on the random treatment assignment to accurately characterise the mNutrition benefits. In addition to the ITT estimates, the quantitative component calculates the local average treatment effects

(LATE) on compliers, which under additional assumptions approximate the impact of receiving the mNutrition messages on outcomes for households that were induced to receive the mNutrition content by the random door-to-door offer ('compliers'). Though the LATE estimates offer an appealing alternative representation of the impacts, they require an accurate measure of exposure to the mNutrition service to be valid. Measuring exposure to the mNutrition programme or use of the service at the household level was challenging because survey respondents often did not know whether SMS messages they received came from the mNutrition service. While the quantitative component was reluctant to rely too heavily on the self-reported measures of service exposure and interpret the LATE estimates as upper bounds of the true LATE impacts for compliers, the LATE treatment effects for compliers provide a potentially more policy-relevant parameter than the previously discussed ITT effects: they represent, albeit for a specific sub-population (compliers), the causal effect of exposure to the mNutrition messaging. For this reason, the cost-effectiveness component of the evaluation considers that the LATE estimates are more representative of the benefit of those who take up the service, and therefore the LATE estimates are used for the effectiveness. The LATE estimates for complier households are between 2.5 and 2.85 times as large as the ITT estimates.

Effectiveness in the quantitative study

The quantitative component found **no evidence that access to the mNutrition service had any impact on child nutrition**, as measured by anthropometry (height-for-age z-scores (HAZ), weight-for-height z-scores (WHZ), stunting, wasting): the ITT analysis found no evidence of a difference between treatment and control groups. Results for the other primary and secondary outcomes are more positive. **The offer of access to the mNutrition service improved dietary diversity and adequacy for children**. Children aged 6–35 months in treatment villages consume from 0.107 more food groups (standard error 0.060) and are 3.8 percentage points (standard error 0.023) more likely to satisfy the minimum dietary diversity threshold. Children aged 6–23 months from treatment villages are also 6.9 percentage points (standard error 0.034) more likely to achieve a minimum acceptable diet. Accordingly, the improved diet diversity for children is taken as the main measurable impact for the cost-effectiveness component of the evaluation. **Women's diets in treatment households are more diverse because of the mNutrition service**. While the impact on the number of food groups consumed is not significantly different from zero, treatment women are 4.0 percentage points (standard error 0.021) more likely to meet the minimum dietary diversity for women threshold, suggesting that they are more likely to have nutritionally adequate diets. The quantitative component observes these dietary improvements for women and children, despite there being no evidence of changes in household-level food consumption patterns. Gilligan *et al.,* (2019) note that '*The main ITT evidence therefore suggests that the mNutrition messaging was an effective way to change beliefs and knowledge and improve the diets of women and young children in the study areas. However, these changes in dietary diversity did not translate into differences in child anthropometry. This indicates that the main barriers to early childhood growth failure in rural Iringa may not be information related, or that changes in nutrition knowledge—on their own—were not enough to affect child nutritional status*.'

User numbers

New registrations fluctuate monthly, as do opt-outs. In the period May 2017 to October 2018 new registrations ran at an average of around 20,000 per month, whereas people left the service at a much lower rate of less than 1,400 per month, i.e. 7% of new users. Numbers of new users, people leaving the service, and the number of users already on the service as at November 2016 (the start

of reliable registration data), combine to give an estimated 550,000 active users as at November 2018.

Effectiveness as DALYs

The cost-effectiveness methodology proposed using the primary outcomes of the quantitative study as the basis of the cost-effectiveness. With the finding that there was no effect in regard to child nutrition, there were no measurable disability-adjusted life years (DALYs) averted. However, to explore some measure of cost-effectiveness, the analysis has been based on estimating the benefit to be derived from an increase in **dietary diversity for 6–35-month-old children** of 0.11 food groups, using relationships published in the literature. The report includes sensitivity analyses to the key assumptions. We acknowledge that trying to link dietary diversity or nutrient adequacy to DALYs could be challenged, particularly in the light of the quantitative study that showed there was no evidence that access to the mNutrition service had any impact on child nutrition. However, in doing so **we make an assumption**: that over the longer term the changes in diet diversity may cause changes in stunting. **The resulting analysis presented here is at risk of being generous to the programme and policy actors need to be aware of that.**

Utilising the Global Burden of Disease data and making assumptions which are outlined in the report, reported dietary diversity changes (using ITT figures) would result in a reduction in the rate of stunting from 36.33% to 36.11% across the nationwide population of children aged 6–35 months and would reduce the number of cases of stunting by 9,440, reducing the burden of disease by 935 DALYs.

Assuming that the composition of the health burden on children reached by Wazazi Nipendeni is similar to that of the health burden on children nationwide, then the reduction in the health burden can be calculated on a pro rata basis. The number of Wazazi Nipendeni users represent roughly 12.8% of the 4.28 million children estimated to be aged 6–35 months. The LATE estimation produced by the quantitative component is a better estimate of the effect found in the 550,000 users. **Given this focus on the treatment take-up, the DALYs averted is stated as 286 DALYs.**

Costs of the service

The baseline report proposed scenarios for the Cost-effective Analysis (CEA), designated Analyses A and B. In Analysis A, it was proposed to include only the ongoing implementation costs and those directly associated with the setting up of the specific service (e.g. the localisation of the content development). This analysis is of primary use to stakeholders who may wish to replicate the service without further development costs, who can implement such a service through an equivalent ownership/partnership with a government, and who has a willing MNO in place. Analysis B was intending to include sunk costs or investments in the project development and the supporting infrastructure. Analysis B is of most use to FCDO and other funders or policy actors to assess whether the programme of mNutrition represents value for money. It would be of particular use if a similar programme was being planned for the future.

Costs were collected from project budgets, expenditure reports, and key stakeholder contacts from multiple organisations as available from inception to July 2019. In the baseline report (Batchelor *et al*., 2018) it was proposed that the relationships between key stakeholders could be regarded as a multi-sided platform business model. This provides a means of making a product free for one group of customers, while another group pays. When considering financial viability, the principal cost components are as follows:

Fixed costs:

- developing content (the Global Content Partners and localising of content);
- mNutrition programme support (e.g. GSMA business intelligence); and
- developing the Wazazi Nipendeni service (mHealth Tanzania Public–Private Partnership (mHealth Tanzania-PPP)).

Operating costs:

- Tanzania mHealth-PPP (offices, staff, servers);
- government agencies (staff costs);
- Short Message Service (SMS) and Unstructured Supplementary Service Data (USSD) costs Mobile Network Operators (MNOs); and
- field partner agencies (staff costs and expenses).

There are three critical assumptions made regarding the costs. The SMS are donated by the MNO, and the real cost of them could be taken as anywhere between a nominal amount (0.003 US cents – Keshav, 2009) and the retail price. They are reported by the PPP as a contribution at retail costs. However, this high SMS price assignment could be challenged. Similarly, the management costs can also be queried. The Cardno reports assigns salaries and wages and fringe benefits in full to Wazazi Nipendeni. However, Cardno manages much more than Wazazi Nipendeni and even within Wazazi Nipendeni, there are activities beyond the mNutrition message sharing. The authors have assigned 50% of the costs of Cardno to the service on the assumption that if the service were replicated without the synergy of all that Cardno does, the staff would have to be in place for Wazazi Nipendeni to get the same or similar reach and to develop the partnerships required with non-governmental organisations (NGOs), government, and MNOs. Section 7 includes a sensitivity analysis on this assumption. Likewise, Rasello also undertakes work other than the mNutrition messaging, and its contract reflects that. Again, the team have assumed 50% of Rasello's contract is assigned to Wazazi Nipendeni. **Taking into account the boundaries of these assumptions the total price can be estimated as between £1,182,000 and £4,792,000.**

For the CEA, we make assumptions at 50% for staffing and the Rasello platform, and assigning a bulk purchase price of 25 TZ Shillings to the price of SMS. The cost of the programme for Analysis A is therefore taken at £2,796,000 for the two years. **For the 550,000 current users (two years of added nutritional content), this suggests the current costs are £5.08 per user (a range of between £2.14 and £8.71 per user).**

The ITT estimates 119 DALYs averted, while the LATE estimates 286 DALYs averted. Taking the LATE estimates as more representative of the effect on users, 286 DALYs suggests **a cost per DALYs averted of £9,776 (a range of £4,132 to £16,755 (2019 prices)).**

Even at the lowest estimate, this exceeds the World Health Organization (WHO) Commission for Macroeconomics and Health (WHO, 2001) guideline for a cost-effective intervention (i.e. <3 times per capita GDP; the GDP of Tanzania is £622 per capita1) (Analysis A).

As discussed, the dietary diversity-stunting-DALYs averted pathway has many assumptions and can be challenged. It is possible that it is underestimating the benefits of the programme, but it is nevertheless one interpretation of the measured results.

[1] 'An intervention is considered very cost-effective, if the monetary amount spent on the intervention per disability-adjusted life year (DALYs) saved is less than the per capita gross domestic product (GDP) for the nation in which the intervention is applied.' 'An intervention is considered (moderately) cost-effective, if the monetary amount spent on the intervention per DALYs saved is less than three times the per capita GDP.'

Commercial sustainability

Having shown that the cost per DALYs for the whole programme is outside the parameters of the WHO guidelines on cost-effectiveness, the report considers the commercial sustainability. Under the current arrangement, the mHealth Tanzania-PPP is funded by public money (from international donors), and the service is enabled by the charitable donation of SMS messages by the MNOs. The costs of developing the Wazazi Nipendeni service have been covered by the US Centers for Disease Control and Prevention (CDC), as part of public health programmes. The costs of developing the additional nutrition content have been covered by the mNutrition programme. **The service is funded and regarded as providing a public good, which it undeniably is. However, the quantitative study has shown that the MNOs are reaping tangible, if indirect, financial benefit from the service.**

In the case of Wazazi Nipendeni it is difficult to argue a case for the financial viability of the mHealth Tanzania-PPP because none of the partners generates revenue directly from providing the Wazazi Nipendeni service. Indeed, all parties are prohibited from generating direct revenue because the Government of Tanzania has ruled that all health services should be free to consumers. **However, the quantitative research has found that average revenue per user (ARPU) increased among users, so it is possible to develop a business case for MNOs to provide an mHealth service on the basis of indirect benefits, should this be the model that is transferred to another country.** If there is a commercial case for MNOs to provide an mHealth service, then there is also a case for a third party to provide content on a revenue share basis (as with conventional value-added services (VASs)).

The quantitative research found that the ARPU was Tanzania shilling (TZS) 510 higher among communities offered Wazazi Nipendeni (compared with control communities). Given an ARPU of TZS 5,300 per month (£1.90/month) among the control sample, this indicates that the VAS stimulates a 10% increase in ARPU. This is supported by findings from the qualitative research that women became more comfortable and confident in using their phone when receiving messages from Wazazi Nipendeni, implying that they then used their phone more, using more airtime. We have made an assumption that the ARPU difference between users and non-users is due to Wazazi Nipendeni.

An analysis of the current arrangement, based on the user growth profile, results in a final contribution margin of 24%. This has been based on retail SMS prices (used by Cardno to report contributions by its donors). However, given that MNOs are sending out messages to thousands of users, it could be argued that a bulk SMS price would be more realistic. A price of TZS 25 per SMS, for example, would give a contribution margin of 70%. If, on the other hand, there is no real cost to the MNO of sending SMS messages (Keshav, 2009), then the contribution margin tends towards 100%. This leads us to consider two scenarios.

Commercial content provider: This scenario speaks to the opportunity for replicating an mHealth agency, delivering a service similar to Wazazi Nipendeni, in another country. At 100% revenue share with MNOs (i.e. all TZS 510/user/month in increased ARPU goes to the content provider), the model suggests it would break even at the beginning of Year 3. However, the rate of return on investment over a four-year time horizon would still be negative. Given the assumed trajectory of growth in user numbers, much greater revenues are generated if the time scales can be extended. **Therefore, at 100% revenue share, the service could generate a positive internal rate of return (IRR) of 4% over a six-year period. However, rate of return is highly sensitive to revenue share: at 90% revenue share (i.e. the MNO retains only 10% of the increase in ARPU), the IRR turns negative.**

MNO in-house service: There is also a hypothetical business case scenario in which an MNO in another country sets up an mHealth service in-house. The key differences from the commercial content provider scenario considered above is that the MNO would retain all of the additional revenue generated, and there would be no real costs associated with the SMS messages sent. If SMS messages and USSD sessions are zero-rated then it would break even in Year 3, although it would not generate enough cash to provide a positive return on investment over a four-year period. **Assuming the same trajectory of increasing users, the large number of users (rising to 1.1 million) would generate enough revenue to provide a 19% IRR in a five-year period. However, it should be noted that Wazazi Nipendeni, as a PPP, has gained users across all four MNOs and it is not clear whether an MNO in-house service would reach as many users.**

Conclusions

On the basis of Analysis A, the cost-effectiveness based on DALYs averted does not fit within the guidelines for an acceptable intervention. We note the many assumptions made; however, even with the most favourable assumptions, the cost per DALYs averted seems outside the scope of the WHO guidelines. However, if the increased ARPU is taken into account, and there were mechanisms for cost recovery the picture could change. MNOs are actually benefitting from their contribution to the programme. The contribution margin is well within their expectations of running a profitable business, even if the SMS were costed at the retail rate.

If the experience of the MNOs is that they get a reasonable contribution margin, and their 'donation' of SMSs is more than matched by the increase in ARPU, then from their point of view there is no loss of opportunity cost in applying the VAS. The idea that the product is commercially viable, and could be replicated commercially, suggests that the narrative of the cost-effectiveness around DALYs averted could be reconsidered by including the commercial returns. MNOs could be persuaded to offer the service not by definition of the number of beneficiaries, but by it being a sound VAS. Whether it then benefits 100 or 1,000 people need not be part of the decision whether to run it or not – it is a viable commercial VAS. The qualitative studies show that the product is appreciated: the perception is that the product or service is valuable. For an MNO it is a viable service to consider offering.

The system is supported by a PPP arrangement, with strong involvement of the Government of Tanzania to ensure contextualised quality-of-health information. In theory, such involvement might not be in place with an MNO or private sector aggregator in the lead. The provision of information to the subscriber must be free, however the indirect benefits suggest that such services can be 'commercially sustainable'. While the original mNutrition programme documents talked about commercial sustainability based on direct revenue generation, the logical framework evolved to talk about imputed benefits to the MNOs. The business modelling report discusses how changes in ARPU can benefit the overall MNO bottom line without actually showing as a direct revenue from a service. Increased ARPU can indeed make the provision of even a free service 'commercially justifiable'.

The remaining challenge is for the revenue, due to indirect benefits, to not only pay for the ongoing costs (by accounting for them in the service budget) but also (at least in part) for the content. Further research is required to identify a mechanism whereby content created using public funds can be paid for through some form of revenue share of the financial benefit yielded by the services that exploit that content.

Table of contents

List of figures and tables

List of abbreviations

ANCOVA	Analysis of Covariance
ARPU	Average revenue per user
CDC	US Centers for Disease Control and Prevention
CEA	Cost-effectiveness analysis
COUNSENUTH	Centre for Counselling, Nutrition and Health Care
DALYs	Disability-adjusted life years
DFID	UK Department for International Development
DHS	Demographic and Health Survey
FCDO	Foreign, Commonwealth & Development Office
GAIN	Global Alliance for Improved Nutrition
GSMA	GSM Association
HAZ	Height-for-age Z-score
HPHB	Healthy Pregnancy, Healthy Baby
ICT	Information and communication technology
IDS	Institute of Development Studies
IFPRI	International Food Policy Research Institute
IRR	Internal rate of return
ITT	Intent to treat
IYCF	Infant and young child feeding
LATE	Local average treatment effect
LiST	Lives Saved Tool
MAD	Minimum acceptable diet
MDD	Minimum dietary diversity
MDD-W	Minimum Dietary Diversity Score for Women
mHealth Tanzania-PPP	mHealth Tanzania Public–Private Partnership
MNO	Mobile network operator
MoHCDGEC	Tanzanian Ministry of Health, Community Development, Gender, Elderly and Children

MOTECH	Mobile technology for community health
NGO	Non-governmental organisation
OLS	Ordinary least squares
OPM	Oxford Policy Management
QTR	Quarter
SMS	Short Message Service
TFNC	Tanzania Food and Nutrition Centre
TTCL	Tanzania Telecommunications Company Limited
TZS	Tanzania shilling
USAID	United States Agency for International Development
USSD	Unstructured Supplementary Service Data
VAS	Value-added service
WASH	Water, sanitation and hygiene
WAZ	Weight-for-age Z-score
WDDS	Women's dietary diversity score
WHZ	Weight-for-height Z-score
WINNN	Working to Improve Nutrition in Northern Nigeria

1 Introduction

mNutrition is a global initiative supported by FCDO, managed by GSMA, and implemented by in-country MNOs and third-party providers that seeks 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, IDS, and IFPRI were contracted to 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.

In Tanzania, the service Wazazi Nipendeni, focuses on the provision of nutrition and health information and services, to vulnerable pregnant women and caregivers of children under the age of five, on their mobile phones, with the goal of improving nutrition outcomes and behaviours for mothers and young children.

1.1 Objectives

The mNutrition evaluation intends to understand and measure the impact, cost-effectiveness, and commercial viability of the mNutrition product using a mixed-methods evaluation design. The evaluation includes a quantitative component, a qualitative component, and a business model analysis. The evaluations were conducted by a consortium of researchers from Gamos, IDS, and IFPRI. The team drew on a number of methods and interlinked workstreams to gather evidence about the impact of the mNutrition intervention in Tanzania:

- the **quantitative impact evaluation**, employing a cluster randomised controlled trial to determine the causal effect of the service;
- a **qualitative impact evaluation**, which consists of three qualitative data collection rounds (i.e. an initial exploratory qualitative study, in-depth case studies at midline, and rapid explanatory qualitative work after the quantitative endline survey data collection), and which aims to provide understanding of the context, underlying mechanisms of change, and the implementation process of mNutrition; and
- a **business model and cost-effectiveness evaluation** employing stakeholder interviews, commercial and end-user data, document analysis, and evidence from the quantitative and qualitative evaluations to generate a business model framework and to estimate the wider imputed benefits from the VAS for the range of stakeholders involved.

The mixed-method evaluation design addresses the following research questions specified in the terms of reference (Annex A):

1. What are the impacts of mobile phone-based nutrition and agriculture services on nutrition, health, and livelihoods outcomes, especially among women, children, and the extreme poor, and how cost-effective are such services?
2. How effective are mobile phone-based services in reaching, increasing the knowledge, and changing the behaviour of the specific target groups?
3. 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 context?

4. What factors make mobile phone-based services effective in promoting and achieving behaviour change (if observed), leading to improved nutrition and livelihood outcomes?
5. How commercially viable are the different business models being employed at country level?
6. 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?

1.2 Research questions for the cost-effectiveness component

The business model and cost-effectiveness component of the evaluation is designed to contribute evidence to help answer the first of the broad research questions specified in the terms of reference (Annex A), and these three questions:

- What are the impacts of mobile phone-based nutrition services on nutrition, health, and livelihoods outcomes, especially among women, children, and the extreme poor, and how cost-effective are such services?
- 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 mNutrition intervention is being externally evaluated in two countries. In Tanzania, where the research consortium is evaluating mNutrition within a broader mHealth service, the intervention aims to promote behaviour change around maternal and early childhood health and nutrition. The target group of the quantitative component was therefore comprised of pregnant women and caregivers of children under the age of five years who reside in rural areas of the study region (Iringa). In Ghana, the intervention is implemented via an mAgriculture programme in which nutrition information has been integrated with crop information as part of a package of agriculture support services. The target group are low-income farmers in rural areas throughout the country. The terms of reference refer to the impacts and effectiveness of mobile phone-based services, so the scope of the evaluation is the mobile-based service as deployed under the mNutrition programme, rather than the incremental impact of support provided through the mNutrition programme. For the quantitative sample of pregnant women and caregivers of young children in rural Iringa that were selected to participate in the quantitative study, the evaluation focuses on estimating the causal effect of access to the mNutrition service. In Tanzania there are no primary livelihood outcomes, there is no disaggregation of the extreme poor due to the way the service is implemented.

The cost-effectiveness component of the evaluation relies on other components in the evaluation study. To determine whether the mNutrition programme in Tanzania is meeting its stated objectives and targets, the quantitative impact evaluation has employed a cluster randomised controlled trial to determine the causal effect of the service. That is, the evaluation seeks to identify how nutrition-related behaviours, knowledge, and outcomes are altered for service beneficiary households relative to their counterfactual levels: what the value of the outcome would have been for beneficiary households in the absence of access to the mNutrition service. In this report we are therefore assessing the cost-effectiveness of Wazazi Nipendeni delivered with the mNutrition content and managed by Cardno and delivered through a platform implemented by Rasello.

While the quantitative evaluation is designed to produce evidence to contribute to the broader research consortium's answers to the first two questions listed in the terms of reference, IFPRI also specified a set of primary and secondary research questions that are answered using information

collected by the quantitative research team. For each of the primary and secondary research questions, the evaluation focuses on estimating the causal impact of the offer of access to the mNutrition service and of registration for the mNutrition content among households induced to participate in the service by the treatment offer. The primary research questions that are addressed through the quantitative evaluation were:

1. What is the impact of the mNutrition service on women's dietary diversity?

2. What is the impact of the mNutrition service on IYCF practices?

3. What is the impact of the mNutrition service on the nutritional status for children under 12 months of age at baseline?

The three primary research questions specify the main outcomes that were studied under the quantitative component of the evaluation. These directly contribute to answering the first overall study question of the mNutrition evaluation (see Section 1.1, above). The impact questions feed into the CEA presented here.

The intended audience for the cost-effectiveness endline report is FCDO, along with other organisations involved in mNutrition and mHealth programmes globally (including local MNOs and NGOs implementing mNutrition services), national governments—in particular, the Tanzanian Ministry of Health, Community Development, Gender, Elderly and Children (MoHCDGEC) and the Tanzania Food and Nutrition Centre (TFNC) in Tanzania—international agencies and donors, and community-level health workers. The report should also be of interest to MNOs who are implementing or considering implementing a similar service.

The reports from the evaluation will be publicly available on IFPRI's and IDS's websites.

1.3 Purpose and scope of the cost-effectiveness endline

This report is the final milestone in the evaluation study; as an endline report, it presents the analysis of the cost-effectiveness of the Wazazi Nipendeni intervention. This report seeks:

- to build on the relevant comparative literature, enabling the cost-effectiveness of the intervention to be compared with programmes that have some similar elements to Wazazi Nipendeni;
- to utilise the framework proposed for the analysis as presented in the baseline report;
- to compare the nutritional impact as defined by the quantitative surveys to the costs, giving an indication of cost-effectiveness; and
- to update the known costs.

In addition, Question 5 of the terms of reference asks for commentary on the commercial viability of the service. While the Tanzania business modelling endline report (Scott *et al.*, 2020) considers this research question in detail, this report also comments on commercial viability in the context of cost-effectiveness.

The report is one of four endline deliverables on the Wazazi Nipendeni service. This report should be read in conjunction with the endline business modelling report (Scott *et al.*, 2020) and the cost-effectiveness baseline report (Batchelor *et al.*, 2018). The quantitative baseline report (Gilligan *et al*., 2018) and qualitative baseline report (Barnett *et al*., 2017) give additional insights into the

consumer environment that the service is targeted at. A final mixed methods report (Barnett *et. al*, 2020) bringing together and summarising the results of the four components is also available .

1.4 Organisation of the report

Section 2 presents a background to the project, while Section 3 discusses the evaluation design. Section 4 looks at effectiveness as captured by the quantitative component of the evaluation. Section 5 considers the measured change with regard to users, to provide insight into the DALYs averted. Section 6 looks at the costs and considers cost-effectiveness Analyses A and B. Section 7 looks at these costs from a commercial sustainability point of view. Section 8 presents a discussion of the findings and Section 9 presents conclusions and learnings.

2 mNutrition intervention

The mNutrition initiative is a five-year global programme supported by FCDO, managed and supported by the GSMA, and implemented by in-country MNOs and third-party organisations that seeks to use mobile technology to improve the health and nutritional status of children and adults in low-income countries around the world. The mNutrition initiative is implemented through existing mAgri and mHealth programmes in 12 countries throughout sub-Saharan Africa and South Asia. The nutrition content aims to promote behaviour change around key farming practices and around dietary and child feeding practices likely to result in improved nutritional health within a household.

In Tanzania, the mNutrition service that is the focus of the evaluation and this report is an integrated programme that enhances an existing SMS-based health communication campaign targeting pregnant women and mothers of young children known as Healthy Pregnancy Healthy Baby. The mass media programme accompanying the service is called Wazazi Nipendeni. The Wazazi Nipendeni programme is a CDC-funded project bringing together multiple partners contributing towards shared goals. Phase 1 of the service, launched in 2012, was initially developed in coordination with the Tanzania Capacity Communication Project, a United States Agency for International Development (USAID) funded service led by the Johns Hopkins Center for Communication Programs. Wazazi Nipendeni is one of several behaviour change services, using methods as diverse as TV drama series, radio distance learning for community health volunteers, and several integrated mass media campaigns. The mass media campaign was developed by the Johns Hopkins Center for Communication Programs, while the SMS component of the campaign is led by the mHealth Tanzania-PPP. The PPP was initiated by the MoHCDGEC, with financial support from CDC. Wazazi Nipendeni is available nationally and on all phone networks.

The HPHB SMS service sent free text messages with healthcare information to pregnant women, mothers with newborns, male supporters, and general information-seekers in Tanzania to drive health-seeking behaviour. The SMS messages are sent in Swahili, originally to women up to 16 weeks post-partum, on a range of pregnancy and early childhood issues timed to coincide with the stage of the pregnancy and the age of the child. Anyone interested in receiving healthy pregnancy information can text the word ‘MTOTO’ (child) to the short code 15001. Registrants receive instructional messages, allowing them to indicate the woman’s current week or month of pregnancy (or the age of the newborn baby) during the enrolment process. This process allows the recipients to receive specific text messages relevant to the time and stage of pregnancy or early childhood. The message frequency also varies depending on the life stage of the woman and child, ranging from nearly daily during key periods of pregnancy to less than weekly for mothers of children over the age of two. Nutrition-related content was a small component of the original HPHB SMS service but was expanded substantially with the addition of the content contributed through GSMA under the mNutrition service. mNutrition added roughly 120 nutrition messages, which are delivered to mothers or caregivers of children up to five years old. During the study period HPHB and mNutrition were available to households in all regions of Tanzania, on all mobile phone networks, and participating individuals received the text messages free of charge. The resulting product is referred to as the mNutrition service in the remaining sections of this report.

The 120 nutrition messages included in the mNutrition service are drawn from 42 factsheets on nutrition-related behaviours identified as key determinants of outcomes that were developed by GAIN together with local partners: the Centre for Counselling, Nutrition and Health Care (COUNSENUTH) and Every1mobile. The information contained in these factsheets was then adapted to the context of Tanzania and made mobile-ready by the local content providers under the guidance of MoHCDGEC and TFNC. As a part of the adaptation process the message content was tested with potential users in Tanzania, after which the language and substance was adjusted

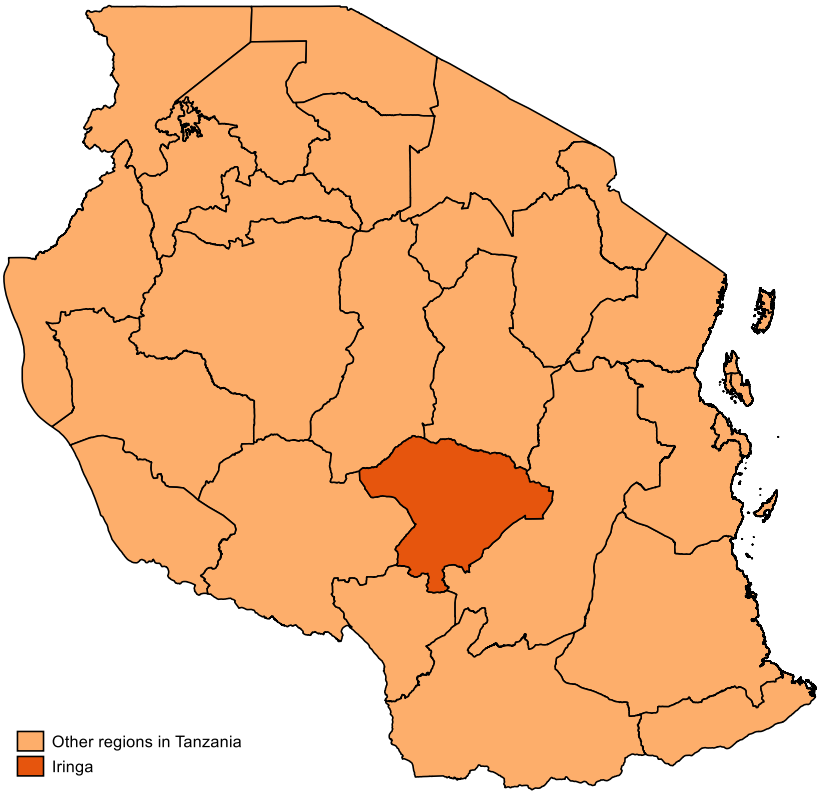
and messages that were identified as not being useful were removed from consideration for the final service. The message testing process highlighted the importance of replacing technical terminology that was likely to be unfamiliar to the message recipients with language that was more commonly used but that still conveyed the evidence-based content of the original factsheets. Included in the final service are messages that encourage the consumption of iron folic acid tablets during pregnancy and messages that promote vitamin A-rich complementary feeding practices and the inclusion of animal source foods in young children's diets, as well as messages providing information on other behaviours accepted as being critical determinants of nutrition outcomes. For more details on the intervention, see the baseline report (Gilligan *et al.,* 2018).

2.1 Study region context

The United Republic of Tanzania is an East African nation with an estimated population of 57.3 million (2016), 68% of whom reside in rural areas (World Bank, 2018). As at 2018, 64% of working-age males and 69% of working-age females were employed in agriculture (International Labour Organization, 2018), with the main agricultural export commodities including tobacco, cashew nuts, coffee, cotton, and sesame seed (Food and Agriculture Organization of the United Nations, 2018). Tanzania is divided into 31 regions and regions are further subdivided into a total of 169 districts. Child undernutrition is a pervasive problem in Tanzania, particularly among young children. In the 2016 Demographic and Health Survey (DHS), 34.4% of children under five were identified as being stunted (HAZ below -2). Wasting is less common, with only 4.4% of measured children under five having a WHZ below the -2 threshold. Over half (57.7%) of measured children between 6 and 59 months of age are anaemic (DHS, 2016).

The quantitative and qualitative components of the evaluation took place within the three rural districts of the Iringa region in Tanzania: Iringa rural, Kilolo, and Mufindi. Figure 1 displays the location of Iringa region in Tanzania. Iringa became an independent region in 1964, before which it was a part of the Southern Highland Province. At the time of the 2012 Population and Housing Census, the total population of Iringa region was estimated to be 0.9 million, 73% of whom resided in a rural area. Agriculture is the primary means of livelihood for most households in the rural parts of the Iringa region: roughly 89% of households in Iringa rural, 92% of households in Mufindi, and 92% of households in Kilolo were involved in agriculture. Average household size was 4.2 and the adult literacy rate was 79% among the rural population (Population and Housing Census, 2012). In Tanzania more broadly, child undernutrition is a severe problem in Iringa. In the 2016 DHS 41.6% of children under the age of five were found to be stunted. This figure was nearly 7 percentage points higher than the national average, suggesting child malnutrition may be more prevalent in Iringa than elsewhere in Tanzania. Additionally, 3.6% of children under five were wasted and 40.3% of children 6–59 months of age were anaemic.

Figure 1: Iringa region in Tanzania



Source: (Gilligan et al. forthcoming, 2020)

Iringa was selected as the location for the study in part because of a dearth of existing relationships between the mHealth Tanzania-PPP and organisations with a presence in the region. Consistent with this, the mNutrition baseline qualitative report found that households in study villages typically rely on health workers at local health clinics and community health workers for their nutrition information needs. Often the information from health clinic workers is received during antenatal visits, which also involve the provision of non-nutrition-related information, testing, and other services, sometimes leaving little time for issues related to nutrition (Barnett *et al.*, 2017). Though at the start of the study there was limited availability of nutrition information and nutrition services, the Government of Tanzania has prioritised improving nutrition outcomes nationwide through different initiatives—for more information, see the baseline report (Gilligan *et al.*, 2018).

2.2 Summary of the quantitative component approach

The quantitative evaluation was designed as a cluster randomised controlled trial, with two stages of randomisation: a village-level randomisation where villages were assigned to a treatment group or to a control group, and a household level randomisation within treatment villages whereby households were either assigned to receive the mNutrition content on just the mobile phone of the primary female or on the mobile phones of both the primary female and the primary male. In villages that were assigned to the treatment group, sampled households were offered access to the mNutrition content on their mobile phone, free of charge, through a door-to-door, in-person visit. In villages that were assigned to the control group, no offer of access to the service was made. Though registration for the mNutrition service was possible for all households regardless of treatment assignment, pre-baseline discussions with the organisation implementing the mNutrition service in Tanzania suggested that take-up of their existing service was low in the study region.

To ensure that the evaluation accurately measures the causal impact of access to the mNutrition service, the quantitative evaluation was based on a cluster randomised controlled trial. From the randomly selected sample of villages participating in the evaluation, IFPRI randomly assigned households in half the villages to a treatment group (T)—where sampled households received a door-to-door offer of access to the content—and households in the other half of participating villages to a control group (C) that did not receive a similar offer. Because the assignment of villages was random, any average difference in outcomes between households in the two groups can be attributed to the difference in access to the mNutrition service.

In addition to the village-level randomisation, the evaluation also included a second-stage household-level randomisation within treatment villages: households in treatment villages where both the mother of the young child or pregnant woman (the primary female) and the primary male own distinct mobile phones (and were surveyed) were randomly allocated to either just receive the mNutrition content on the mobile phone of the primary female (T-F), or to receive the mNutrition content on the mobile phone of the primary female and the mobile phone of the primary male (T-F+M). By comparing behaviours and outcomes between treatment households in the T-F+M group and those in the T-F group, and contrasting both the T-F and the T-F+M group to households in control villages that would have been eligible for the household-level randomisation, it was expected that learning would be generated about how information flows between spouses.

Table 1: Summary of household surveys in the treatment and control group for Quant survey

	Treatment	Control	Total
Number of EAs completed	90	90	180
Baseline household interviews	1,428	1,405	2,833
Attempted endline household interviews	1,428	1,405	2,833
Completed endline interviews	1,243	1,234	2,477
Partially completed	51	67	118
Long term unavailable	57	45	102
Household not found	10	2	12
Household located elsewhere	65	57	122
Refusal	2	0	2

Source: Gilligan *et al* (forthcoming, 2020)

2.3 Summary of the qualitative component approach

The sample selection for the qualitative study was purposive and based on the quantitative baseline data. Three sites were selected from the sample of 90 treatment villages. At community level, the main data collection tools were semi-structured in-depth interviews with treatment mothers and fathers (i.e. mothers and fathers who were signed up to receive Wazazi Nipendeni plus mNutrition messages by the OPM Tanzania team during the quantitative baseline survey), key informant interviews, and focus group discussions with treatment mothers and fathers and elderly women. In-depth interviews were conducted with mothers who had signed up to receive the Wazazi Nipendeni mNutrition service on their phone or their partners' mobile phone as part of the baseline survey. In-depth interviews were also conducted with fathers whose partners were signed up to the service, who received messages on behalf of or in addition to their wives. Key informant interviews were conducted with influential and knowledgeable people in each village, including village chairmen and/or community health workers. The purpose was to explore their knowledge, awareness, and opinions of the mNutrition messages that people in the community had been

receiving, and to capture information on current contextual issues relating to maternal and child nutrition. Focus group discussions were carried out with four or five participants per group. The aim was to select heterogeneous groups of mothers (by occupation, age, number of children) who had signed up to the Wazazi Nipendeni plus mNutrition service.

3 Evaluation design

3.1 Study design

The aim of the impact evaluation is to assess the impact, cost-effectiveness, and commercial viability of mNutrition. The evaluation is being conducted by a consortium of researchers from Gamos, IDS, and IFPRI. The team is drawing on a number of methods and interlinked components to gather evidence about the impact of the mNutrition intervention in Tanzania, including a qualitative component, a quantitative component, and a business model and cost-effectiveness component.

This report focuses on the cost-effectiveness of the proposition.

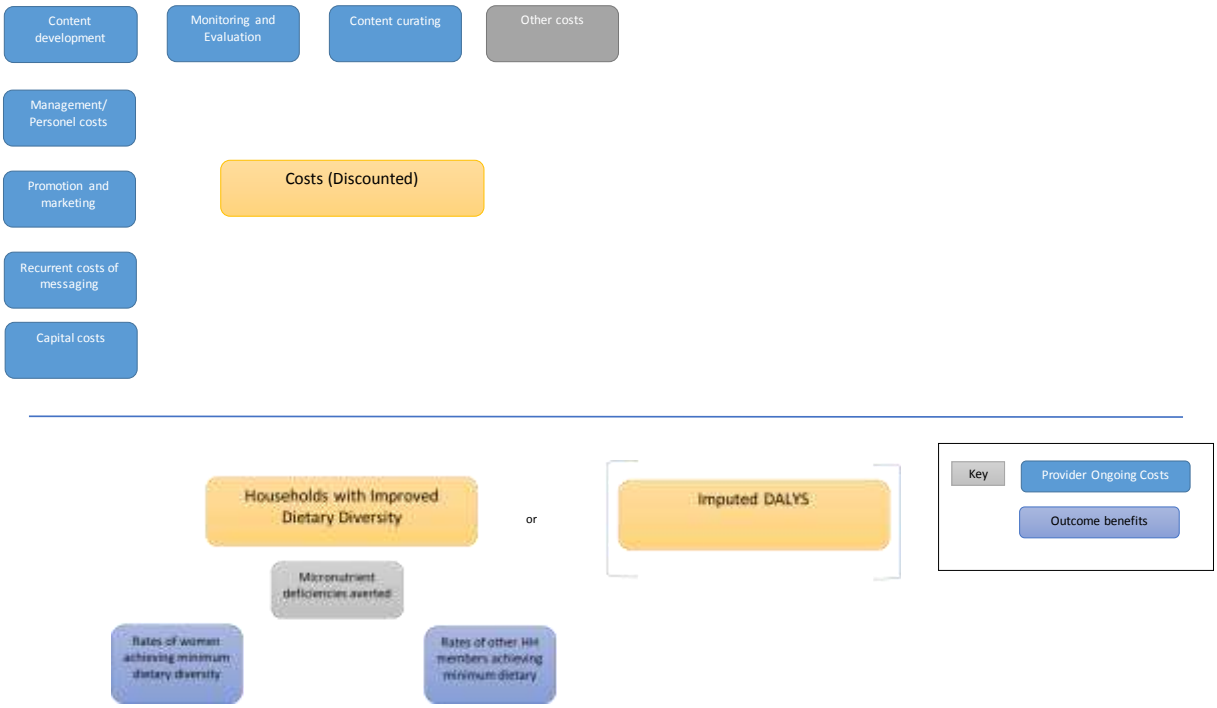
The baseline report proposed three scenarios for the CEA, designated Analyses A, B, and C.

In Analysis A, it was proposed that above-the-line costs set out in Figure 2 include only the ongoing implementation costs and those directly associated with the setting up of the specific service (e.g. the localisation of the content development). Analysis A does not attempt to allocate a proportion of the wider mNutrition programme costs, nor take into account the sunk and investment costs associated with building the asset value of the PPP linkages across other health campaigns, or the network infrastructure brought by each participating MNO.

Below the line, the effectiveness was to be assessed by reference to the measured change in anthropometry and/or dietary diversity of mother and child. This has been measured within the quantitative component of the overall study.

This analysis is of primary use to stakeholders who may wish to replicate the service without further development costs, who can implement such a service through an equivalent ownership/partnership with a government, and who has a willing MNO in place. It is based on the assumption that content, at least at the global level, will be available free of charge. This is indeed the case, given that all the factsheets and messages developed by the Global Content Partners for the mNutrition programme are open access and have been made freely available through the CABI Knowledge Bank. It also assumes that any future implementing agency will have access to the technical platform and capability needed to implement such a system, either as part of their own resources, or by subcontracting the services of a company that does have such capability, such as Rasello.

Figure 2: Costs associated with Analysis A



Source: Batchelor et al 2018

Analysis B (illustrated in the baseline report) was intending to include sunk costs or investments in the project development and the supporting infrastructure. The wider costs were difficult to apportion. One could argue that a portion of the wider nutritional project research and development should be assigned to the Wazazi Nipendeni costs, since if donors were to invest in, for example, a next generation of mNutrition services, they would need an overarching programme of work similar to the mNutrition programme to stimulate MNOs to adopt new approaches, to coordinate learnings and effort, and to deliver wider programmatic benefits. The same argument applies to the global content developed as part of the overall mNutrition programme, and to the institutional infrastructure set up by GSMA (and others) in order to deliver the programme.

Analysis B is of most use to FCDO and other funders or policy actors to assess whether the programme of mNutrition represents value for money. It would be of particular use if a similar programme was being planned for the future.

In the baseline, Analysis C was planned in order to take a comprehensive view of costs, taking into account not only operational costs and wider mNutrition programmatic costs, but also wider societal costs. For instance, there is an argument that the real cost of the text messaging service should include sunk investment in the MNOs' operations. In the case of Wazazi Nipendeni, there is also a wider PPP arrangement that facilitates the smaller mNutrition team, which includes the history of information provision and a wealth of learning on what works and what does not – the cost of that learning being sunk costs from previous projects (existing assets). An element of the societal costs could also consider any cost associated with any behaviours that mothers would need to undertake in order to realise the benefits pursued by the project. For example, if the project suggested greater attendance at health clinics, Analysis C might have considered the transport cost to the household of getting to the clinic. However, it was not possible to gather sufficient data to undertake this analysis. Operational costs of the MNOs were not available beyond their direct contribution to the project, the historical investment of donors in the PPP was not available, and the

quantitative data suggested no significant change in key behaviours, such as antenatal clinic attendance. For these reasons, Analysis C was not attempted.

In terms of the perspective of the three analyses, based on Dunet (2012) and Neumann (2009) we suggest the analyses can be considered as shown in Table 2.

Table 2: Perspectives used within the cost-effectiveness analysis

Cost	Perspective			
	Societal	Donor/Govt	Donor/GSMA	PPP/GSMA
Direct Wazazi Nipendeni costs to run the service (Analysis A)	x	x	x	x
Indirect Wazazi Nipendeni costs (Analysis B), including wider GSMA project infrastructure	x	x	x	
Indirect health infrastructure and health-seeking effects and expenses (Analysis C)	x	x		
Intangible, qualitative user opinions	x			

Source: Authors' own

However, as discussed above, Analysis C and a full societal perspective have not been possible.

3.2 Data collection methods

This report has been drafted drawing on interviews with key stakeholders and secondary data. A list of contacts made during the endline collection can be found in Annex B.

- qualitative interviews conducted with stakeholders and MNOs in Tanzania;
- commercial data provided by stakeholders and MNOs (or brokered by GSMA);
- monitoring data gathered by Altai;
- data available in published literature;
- government stakeholders and alternative service providers as a source of additional, unpublished information on costs and business models;
- the quantitative component of the study, led by IFPRI (Gilligan *et al*., forthcoming, 2020); and
- the qualitative component, which focused on consumer perceptions, led by IDS (Barnett *et al*., 2019).

This report draws heavily on data analysis presented in the quantitative endline report (Gilligan *et al*., 2019). Full details of the sampling design and the analytical approach can be found in that report.

3.2.1 Processing information

Evaluation activities carried out by Gamos to inform the endline reports include the following:

- Field visits to establish and maintain relationships with key stakeholders. Interviews conducted with key representatives of stakeholder institutions to gather additional data to populate the CEA framework. Ongoing communication and field visits undertaken to monitor developments in services and to track the commercial justification for changes.
- Working with IDS and IFPRI to contribute to the design of both qualitative and quantitative instruments (both baseline and endline) to incorporate indicators relating to non-financial

attitudes of customers to services, and to MNOs in particular, such as customer satisfaction and brand loyalty. These instruments also explore attitudes towards alternative services offered by other providers, e.g. media, face-to-face extension.

- Analysing the impact data from the quantitative report, and utilising the Global Burden of Disease, Lives Saved Tool (LiST) models, and other related literature and databases for Tanzania to convert the impact to DALYs.
- Analysing financial data with a view to creating a financial model to test key cost sensitivities.

The process of enquiry and information collection was flexible and responsive to events on the ground. This component of the evaluation is based on opportunistic data gathering from key individuals, such as representatives of the core partners and other partners to the project.

3.3 Ethical considerations and approval

As an overall guiding principle, the research team sought to conduct themselves in a professional and ethical manner throughout the baseline phase of work, with strict respect for the principles of integrity, honesty, confidentiality, voluntary participation, impartiality, and the avoidance of personal risk. These principles were informed by the Organisation for Economic Co-operation and Development (2010) Development Assistance Committee Quality Standards for Development Evaluation and FCDO's 'Ethics Principles for Research and Evaluation', which were followed for the duration of the evaluation.

Overall, this component draws on the qualitative and quantitative data collected in the other two components of the evaluation. Other data sources include stakeholder interviews with MNOs and data collection (commercial and monitoring data) from MNOs and other relevant organisations.

Although most research participants were familiar with the mNutrition programme, and with the principle of an independent evaluation, this component sought to obtain the informed consent of participants. This was achieved by emails and briefing documents describing the research. In particular, these described the relationship between the consortium, FCDO, and GSMA, in order to avoid any possibility of deception. Research activities with participants involved interviews only; there were no observational activities.

While this evaluation cost-effectiveness component did not involve any primary data collection from human subjects at community/household level, ethical considerations are still considered important for all work carried out under this component. In particular, GSMA remain highly aware of the commercial sensitivities of their partner MNOs, so the issue of commercial confidentiality is very important for this area of work, given that it relies on the sharing of sensitive commercial data. Therefore, the Gamos team has paid specific attention to this issue as part of their ongoing work.

The Gamos team is currently operating under the non-disclosure agreement signed by GSMA and OPM during the inception phase of the project. Where relevant, stakeholder respondents were informed that a non-disclosure agreement with their trade association had been signed, and that the interview was bound by it. All the data that were gathered fell within the scope of this agreement (e.g. development, business plans, marketing, operations, and finances), although there is a provision that such information should be designated as proprietary or confidential[2].

[2] The agreement permits Gamos to share confidential information among the team if: 1. they need to know; 2. they have entered into a confidentiality agreement; and 3. they are not a competitor.

For the avoidance of doubt, all internal reports shared by Gamos were marked as confidential and were not to be circulated outside of the evaluation team. Any outside reporting will not contain any detail that could be construed as proprietary or confidential information.

All external reports were and will be shared with key research participants in early draft form in order to establish principles of trust and reciprocity. This is to ensure that participants have an opportunity to confirm that their views are reported accurately, and that publications do not breach their confidentiality requirements.

As this component draws on qualitative and quantitative data collected through the other two workstreams, appropriate measures were taken to ensure that the shared data are anonymised and there is no risk of confidentiality breach. For the quantitative data, a unique household ID has been assigned to each household, which allows for following up with respondents as necessary without providing access to any personal information on datasets that are made available for analysis. Similarly, all qualitative transcripts are anonymised and pseudonyms given, and any information that can lead to personal identification has been removed.

3.4 Limitations

The methodology relied on the willingness of key stakeholders to share their data and their thoughts. In a commercial environment this is not always forthcoming, and a limitation of the report is that it relies on these shared data. Risks associated with this transfer of data have been mitigated as much as possible by clear communication and follow-up with stakeholders. The degree of engagement to date with stakeholders is reflected in the insights and level of data presented in the report. Changes in relationships and personnel (among all stakeholders) were the principal threats to the mitigation strategy.

The risks associated with the evolving nature of the business model have been mitigated as much as possible by setting milestone data points, and the subsequent phases have informed the changes between baseline and endline.

The cost-effectiveness is limited by the allocations of costs and the assumptions made surrounding the documenting and modelling of costs.

4 Impact measurement

In this section, we present the impact of the nutrition-enhanced service of Wazazi Nipendeni on primary outcomes identified as measured by the quantitative component, which provides the basis for the analysis of effectiveness within the CEA. These findings are drawn from the quantitative component report (Gilligan *et al*., forthcoming, 2020) and more detail can be found in that report.

4.1 ITT estimates

Because the offer of treatment was randomly assigned, Gilligan *et al*., use the systematic variation in take-up of the mNutrition service generated by the random offers to identify the causal impact of the service. The random assignment of the treatment offers ensures that unbiased estimates of the treatment effects can be estimated using simple differences, difference-in-differences, or analysis of covariance (ANCOVA) specifications because observable and unobservable characteristics of children, households, and communities are balanced across treatment and control villages. However, ANCOVA models, which control flexibly for a baseline measure of the outcome, are likely to be the most efficient of the three estimators, particularly when the autocorrelation for the outcome being considered is low (McKenzie, 2012). Therefore, for panel outcomes—those that are observed at both baseline and endline—they use ANCOVA to generate our primary estimates, and simple differences and difference-in-differences models as robustness checks. In addition to providing more efficient estimates of the treatment effects, the ANCOVA model allows us to estimate the relationship between the baseline and endline measures of the outcome. For outcomes that are observable at both baseline and endline, the quantitative component relies on this ANCOVA specification to generate the main treatment effect estimates.

The quantitative surveys show that the random offer of access to the mNutrition content generated a gap in self-reported participation in the service during the study period. Households in treatment villages were 39.7 percentage points more likely to report having received automatic text messages about nutrition during the two years preceding the endline survey. Even among households that self-identified as service participants in the endline survey, receipt of the mNutrition content was relatively infrequent. This may in part be due to intended variation in message spacing, whereby households were to receive fewer messages as children got older. Other potential drivers of infrequent messaging are mobile phone loss or intentional switching, which could have prevented some treatment households from continuing to receive the mNutrition messaging. Another possibility is that the mNutrition content was received on the mobile phones of some treatment households, but they did not notice the messages on their mobile. These barriers to consistent service exposure are potentially critical for the success of mobile phone-based information interventions, and future research should explore ways to improve mobile phone attachment and message salience to improve the effectiveness of similar programmes.

Another potential barrier to the effectiveness of the mNutrition service is whether households read the messages and are satisfied enough with the message content to implement the suggestions they contain. 85.8% of females who received an mNutrition message and 86.1% of males reported having read all of the content they were sent through the service. Satisfaction with the content was similarly high, with 91.4% of females and 93.2% of males finding the messages they read always useful. This clearly indicates that the mNutrition content was well suited for the households included in the study sample, and that there was interest in the topics and information the service covered. In addition to the primary and secondary outcomes, men and women in treatment villages interacted significantly more with their mobile phones and were more likely to identify automatic text messages as a source of nutrition information. Women were more likely to call, receive calls,

text, receive texts, and receive mobile money if they were in a village randomly assigned to the treatment group; men were more likely to send and receive text messages and to send mobile money. Consistent with the self-reports on increased mobile phone use, the quantitative component found that treatment households spent 9.4% more on mobile phone vouchers than control households. These observed impacts on use of mobile phones suggest that, in addition to generating behavioural benefits for participating households, there may be benefits that accrue to the MNOs from providing the mNutrition service.

4.1.1 LATE for compliers

The specifications described above enable the quantitative study to estimate ITT treatment effects: that is, the point estimates capture the impact of the random offer of access to the mNutrition service on outcomes. However, under two assumptions[3] they also estimate the LATE of receiving the mNutrition content for compliers: households that were induced to register for the service by the randomly assigned door-to-door offer. The first assumption required is that the mNutrition offer only affected outcomes indirectly, by increasing the likelihood that households received the mNutrition content on a mobile phone. As long as the in-person visit and offer of access to the mNutrition service did not directly affect the outcomes of interest—that is, if it did not change child anthropometry, IYCF knowledge, or IYCF behaviours through a pathway other than increased exposure to the mNutrition content—then this assumption will be satisfied. The second assumption necessary for estimating LATE for compliers is that the randomly assigned offer of access to the mNutrition service does not decrease the likelihood that any household or household member actually goes on to register to receive the content.[4] Both of these assumptions are likely to be satisfied in the context of this evaluation (see full discussion in Gilligan *et al*., 2018).

The LATE treatment effects for compliers provide a potentially more policy-relevant parameter than the previously discussed ITT effects: they represent, albeit for a specific sub-population (compliers), the causal effect of exposure to the mNutrition messaging. However, the quantitative team concluded that they prefer the ITT estimates presented to the LATE estimates shown as the best way to characterise the service benefits (Gilligan *et al.*, forthcoming, 2020). In this CEA we have considered the LATE estimates to inform a sensitivity range for the cost-effectiveness.

4.2 Outcome indices

The primary and secondary outcomes related to IYCF practices and knowledge are critical for evaluating the success of the mNutrition service. Because IYCF practices and knowledge are multi-dimensional and include observations and subject responses to a series of questions, the quantitative component condenses the individual responses into indexes, which it calls Anderson indexes. For IYCF knowledge, the authors calculate the percentage of the 11 baseline and 20 endline IYCF knowledge questions that were answered correctly by each female and male respondent. In addition, to generate a measure of household IYCF knowledge, the authors use a method proposed by Anderson (2008) to combine the female and male scores for each household.

The quantitative team used the same method to construct two IYCF practice index scores, one for children 6–23 months of age and one that includes children 6–35 months of age. The IYCF practice indexes combine information about whether children born in the past 24 months were breastfed during the first hour after birth, whether children 0–5 months are exclusively breastfed,

[3] See Imbens and Rubin (2015) for a complete discussion.
[4] Note that this does not require it to be true that treatment households would not have registered for the service if they had not received the door-to-door offer, just that treatment households were more likely to have registered for the service because they received the door-to-door offer than they would have been if they had not received the offer.

whether children 12–15 months are still being breastfed, whether children 6–8 months receive solid, semi-solid, or soft foods, whether children in the relevant age range (either 6–23 months or 6–35 months) consumed food from at least four of seven possible food categories in the 24 hours preceding the survey, and whether children meet the minimum meal frequency given their age and breastfeeding status. In contrast with the IYCF knowledge index described above, no household has non-missing data for all these indicators and several of the indicators are nearly always missing for the households in the sample, due to age restrictions. The authors therefore retain households that have missing values for the component indicators. This has the result of attenuating the size of the IYCF practices Anderson indices, as most households receive zeros for the highly age-specific measures. As mentioned above, the indices will weight these indicators less heavily, as they will have little variance relative to the more populated indicators.

4.3 Primary outcomes: ITT analysis

4.3.1 Child anthropometry

Across all the child anthropometry outcomes, the quantitative component found no evidence that providing households access to the mNutrition service had an impact—positive or negative—on any of the outcomes. The point estimates are nearly identical and in no case are any of the conclusions different when using the two methods. It should be noted that change in the anthropometric measures was not expected and its measurement was based on contractual diligence.

4.3.2 IYCF practices

Table 3 turns to exploring whether the mNutrition service had any impact on IYCF practices as reported by the primary female respondent. Because some of the IYCF practice indicators are only relevant for a narrow range of child ages, that frequently have little overlap with the age at endline of the focus children in our sample, **the quantitative team caution against placing too much weight on the results or lack of significant results for these outcomes.** The dietary diversity indicators—both the count of the number of categories consumed in the previous 24 hours and the indicator for whether children consumed from at least four categories during the previous 24 hours—are available for a larger set of sample children. Additionally, these indicators have been validated as predictors of micronutrient adequacy for 6–23-month-old children and 6–35-month-old children (Arimond *et al.,* 2010; Moursi *et al.,* 2008).

Though positively signed, the quantitative component found no statistically significant impact on the likelihood of early initiation of breastfeeding (point estimate 0.010 with a standard error 0.072 and p-value 0.41 in the basic controls case). Further, at the 5% level the authors can reject modest sized effects larger than a 5.9 percentage point increase or a 3.9 percentage point decrease. The estimates of the impact on exclusive breastfeeding for children 0–5 months and the consumption of solid, semi-solid, or soft foods for children 6–8 months are negative—but they are based on only 165 observations and 58 observations, respectively, and are not statistically distinguishable from zero.

The quantitative component did observe a positive impact on the likelihood of continued breastfeeding for children 12–15 months of age, though there are only 27 children in the sample within this age range. The point estimate suggests that the offer of access to mNutrition increased the likelihood that children 12–15 months of age were still breastfed by 26.2 percentage points (standard error 0.131 p-value: 0.057), or roughly 34% of the control group mean. As mentioned

before, the authors are hesitant to place too much weight on this result as it is driven by there being two control group children 12–15 months who were not breastfed, relative to no treatment group children in that age group not being breastfed.

There are, importantly, observations for the count of the number of food categories children 6–23 months of age consumed and children 6–35 months of age consumed in the 24 hours preceding the survey (843 and 2,558). The quantitative component found evidence that household access to mNutrition had a positive impact on dietary diversity for both age groups, though only the effect for 6–35-month old children is statistically significantly different from zero. Access to mNutrition increased the number of food categories consumed by 0.107 (standard error 0.060 p-value 0.073) in the basic model and 0.112 (standard error 0.056 p-value 0.046) in an extended model. While relative to the control group mean these effects are modest in size (roughly 3% of the control group mean), they are still over half the size of the impact found for a much more intensive behaviour change communication programme in Ethiopia that included in-person counselling (Kim *et al*., 2019), as well as more than one-third of the size of an integrated health and nutrition programme that included food rations which was implemented in Burundi (Leroy *et al*., 2015).

The quantitative component found some evidence that child dietary diversity may be increasing through an increased likelihood of children consuming vitamin A-rich fruits and vegetables or other fruits and vegetables: with point estimates of 0.20 (standard error 0.012 p-value 0.107) and 0.38 (standard error 0.022 p-value 0.087) in the basic models. However, the statistical significance of the service impacts depends on whether the quantitative study uses the basic or the extended model, suggesting these effects are not robust. The authors also found a positive, albeit never statistically significant, impact on dairy consumption, with the point estimate indicating access to the mNutrition service was associated with a 3.0 percentage point increase (standard error 0.020 p-value 0.147) in the likelihood that children consumed dairy in the 24 hours before the survey.

Access to the mNutrition service also appears to have increased the likelihood that children aged 6–35 months met the minimum dietary diversity (MDD) standard (WHO, 2008). Children in treatment villages were 3.8 percentage points more likely to have consumed from four food groups (standard error 0.023 p-value 0.093). This represents an increase equivalent to 6.3% of the endline control group mean and the estimate is unaffected by whether the quantitative study uses the basic or extended model. The corresponding estimate for children aged 6–23 months is much smaller in magnitude (0.001 in the basic model) and not statistically distinguishable from zero. For children aged 6–23 months, household access to the mNutrition messaging was not statistically significantly associated with the likelihood that they met the minimum meal frequency requirements, though the point estimate is positive (coefficient 0.025; p-value 0.343).

Minimum acceptable diet (MAD), the intersection of minimum meal frequency and MDD for children aged 6–23 months, was relatively rare among the control children: just 22.4% of the control group satisfied the MAD requirements. Children in the mNutrition treatment group were 6.9 percentage points (standard error 0.034) more likely to satisfy MAD based on the model with basic controls and 7.8 percentage points more likely to satisfy MAD in the model with extended controls (standard error 0.033). Both estimates are statistically significant at the 5% level, with p-values of 0.042 and 0.020, and the point estimates correspond to increases in the likelihood of satisfying MAD equivalent to 30.8% and 34.8% of the control mean. These are therefore meaningfully large impacts on a MAD indicator that has been associated with both energy intake and mean nutrient density adequacy for vitamins A, B6, C, riboflavin, thiamine, folate, iron, zinc, and calcium (Dewey *et al*., 2006).

The last two rows in Table 3 present the impact estimates for the IYCF Anderson indices for children aged 6–23 months and children aged 6–35 months. For neither index is the estimate of the impact of access to mNutrition statistically significant from zero (p-values of 0.256 and 0.213 in

the basic models and 0.190 and 0.171 in the extended model), though both point estimates are positive, regardless of which model is used.

Table 3: Impact estimates of mNutrition on IYCF practices, simple differences

	Control mean	Impact estimates, basic controls	Impact estimates, extended controls	N
Children born in last 24 months who were put to the breast within one hour	0.776	0.010	0.012	1,011
		(0.025)	(0.025)	
Infants 0–5 months of age who are fed exclusively with breast milk	0.032	-0.014	-0.016	165
		(0.027)	(0.024)	
Children 12–15 months of age who are fed breast milk	0.846	0.262*	0.302*	27
		(0.131)	(0.168)	
Infants 6–8 months of age who receive solid, semi-solid, or soft foods	0.697	-0.050	0.045	58
		(0.130)	(0.151)	
Number of food groups (out of seven) children 6–23 months of age consume	3.685	0.056	0.055	843
		(0.093)	(0.089)	
Number of food groups (out of seven) children 6–35 months of age consume	3.766	0.107*	0.112**	2,558
		(0.060)	(0.056)	
Child consumed grains, roots, or tubers	0.914	-0.002	-0.001	2,813
		(0.008)	(0.008)	
Child consumed legumes or nuts	0.496	0.008	0.013	2,811
		(0.025)	(0.024)	
Child consumed dairy	0.198	0.030	0.028	2,813
		(0.020)	(0.019)	
Child consumed fish or meat	0.442	-0.002	0.003	2,811
		(0.023)	(0.022)	
Child consumed eggs	0.080	0.006	0.006	2,811
		(0.011)	(0.011)	
Child consumed vitamin A-rich fruits or vegetables	0.849	0.020	0.020*	2,813
		(0.012)	(0.012)	
Child consumed other fruits or vegetables	0.579	0.038*	0.035	2,811
		(0.022)	(0.022)	
Children 6–23 months of age who consume four or more food groups	0.609	0.001	-0.001	843
		(0.034)	(0.033)	
Children 6–35 months of age who consume four or more food groups	0.604	0.038*	0.038*	2,558
		(0.023)	(0.021)	
Children 6–23 months of age who meet the minimum meal frequency	0.790	0.025	0.029	846
		(0.026)	(0.025)	
Children 6–23 months of age who meet the MAD	0.224	0.069**	0.078**	830
		(0.034)	(0.033)	
Anderson index: IYCF practices (6–23 months)	-0.003	0.007	0.010	1,011
		(0.007)	(0.007)	
Anderson index: IYCF practices (6–35 months)	-0.001	0.003	0.004	2,734
		(0.002)	(0.003)	

Notes: Estimates from the mNutrition Tanzania endline survey sample. Standard errors are in parentheses and clustered at the village level. Impact estimates report the coefficient on treatment from an ordinary least squares (OLS) regression of the outcome of interest on treatment variable, controlling for baseline household classification, child's age, and child's gender. Extended controls are covariates from baseline: household size, whether household head is female, whether household head is literate in Swahili, whether primary female owns a mobile phone, Progress out of Poverty Index score. Control mean is comparison group's mean at endline. * p<0.10 **p<0.05 ***p<0.01.

Source: Authors' own

4.3.3 Dietary diversity

The quantitative component presented treatment effect estimates for two measures of women's dietary diversity, as set out in Table 4, the Women's Dietary Diversity Score (WDDS), which ranges from 0 to 10 depending on the number of categories women consumed from in the day preceding the survey, and the Minimum Dietary Diversity Score for Women (MDD-W), which is an indicator equal to 1 if women consumed from at least five of the 10 categories. Though it is not a primary or secondary outcome, the authors also show results for the household dietary diversity score as it helps to provide context for the two variables measuring children's dietary diversity shown in Table 3 and the two variables measuring women's dietary diversity shown in Table 4.

The point estimates for both measures of women's dietary diversity are suggestive of the idea that access to mNutrition improved the quality of women's diets. Access to mNutrition increases WDDS by 0.073 in the basic model (standard error 0.068 p-value 0.279) and 0.076 in the extended model (standard error 0.066 p-value 0.249), though neither impact is statistically distinguishable from zero at the 10% level. For MDD-W, mNutrition increases the likelihood that the primary females satisfy MDD-W by 4.0 percentage points in the basic model (standard error 0.021 p-value 0.062) and 3.6 percentage points in the extended model (standard error 0.020 p-value 0.076); the effect sizes for MDD-W represent increases of 6.4 and 5.8% of the control group mean.

There is therefore evidence that mNutrition increased dietary quality for both women of reproductive age and children, and that these increases were focused on individuals at or below thresholds that have been linked to dietary adequacy for these groups (MDD and MDD-W) (p-value <0.1). Interestingly, the authors find no evidence that the mNutrition service had any impact on household dietary diversity. The point estimate, while positive, is small in magnitude (0.020 in the basic model and 0.045 in the extended model with standard errors 0.088 and 0.028 respectively) relative to the control group mean and the corresponding coefficient for women, and not approaching statistical significance at conventional levels (p-value of 0.819 in the basic model and 0.585 in the extended model). Interestingly, this indicates that the improvements in women's and children's diets may have been driven by changes in the way that households allocate food across household members, rather than through affecting household-level consumption patterns more broadly.

Table 4: Impact estimates of mNutrition on dietary diversity, ANCOVA

	Control mean	Impact estimates, basic controls	Impact estimates, extended controls	N
WDDS (0-10)	4.437	0.073	0.076	2,535
		(0.068)	(0.066)	
Met MDD-W	0.625	0.040*	0.036*	2,546
		(0.021)	(0.020)	
Household Dietary Diversity Score (0-12)	5.336	0.020	0.045	2,585
		(0.088)	(0.082)	

Notes: Estimates from the mNutrition Tanzania endline survey sample. Standard errors are in parentheses and clustered at the village level. Impact estimates report the coefficient on treatment from an OLS regression of the outcome of interest on treatment variable, controlling for baseline household classification and value of the respective outcome at baseline. Extended controls are covariates from baseline: household size, whether household head is female, whether household head is literate in Swahili, whether primary female owns a mobile phone, Progress out of Poverty Index score. Control mean is comparison group's mean at endline. * p<0.10 **p<0.05 ***p<0.01.

Source: Gilligan *et al* (forthcoming, 2020)

4.3.4 ITT summary

The quantitative component found no evidence to suggest that access to the mNutrition service had any impact on child nutrition as measured by anthropometry. Point estimates in both the basic model with minimal controls and the extended model with more detailed controls for baseline socioeconomic characteristics result in precisely estimated zeros for the point estimates.

The results for the other primary and secondary outcomes are much more positive. There is evidence that access to the mNutrition service improved dietary diversity for children: both children's dietary diversity scores and the likelihood that children satisfied the MDD threshold were statistically significantly increased by the mNutrition service. The quantitative component also found that mNutrition importantly improved the likelihood that children aged 6–23 months achieved a MAD. Similarly, women's diets in treatment households were also more diverse. Though not statistically significant, treatment women consumed from more food groups, on average, than their control group counterparts. Importantly, this translates into an increased likelihood that treatment women met the MDD for women threshold, suggesting that treatment women were more likely to have nutritionally adequate diets. The quantitative component observed these dietary improvements for women and children despite there being no evidence of changes in household-level food consumption patterns: mNutrition had no qualitatively or statistically significant impact on the number of food groups that households consumed from in the 24 hours preceding the survey.

As expected, given the observed impacts on dietary diversity, mNutrition also improved nutrition knowledge in treatment households. Though the quantitative component estimated a positive and statistically insignificant effect on female knowledge, treatment group males answered 1.7% more IYCF and nutrition knowledge assessment questions correctly than control group males. Similarly, the combined household-level IYCF knowledge score was statistically significantly increased by mNutrition access, with treatment households scoring 0.09 standard deviations higher on the endline knowledge assessment.

Together, this evidence suggests that the mNutrition messaging was an effective way to improve the diets of women and young children in the study areas. The messaging had large impacts on the nutrition knowledge of male household members, and these changes in beliefs were also

associated with changes in how children and women consumed food, specifically increasing the number of food groups these populations consumed from. That these changes in dietary diversity did not translate into decreases in stunting or improvements in HAZ for children likely suggests that the main barriers to early childhood growth failure in rural Iringa are not information- or knowledge-related, or that the changes in knowledge were not, on their own, enough to affect child nutritional status, or that it takes time for the effect of the improved diet diversity to show up as anthropometrics.

Primary outcomes: LATE analysis: The quantitative component extends the main mNutrition impact estimates that were presented in the previous section by estimating the LATE for households induced to receive the mNutrition messages by the randomly assigned offer of access to the service (compliers). **The quantitative component preferred the ITT estimates presented in the previous section as they are policy-relevant, easy to interpret, and require no assumptions beyond the conditional independence assumption that randomisation ensures is satisfied. However, the authors of the quantitative component also recognise that the ITT estimates may understate the impact of the mNutrition service if there is imperfect compliance with the randomly assigned offer of access to the mNutrition content.** If an important share of control households registered for mNutrition during the study period, or if some treatment households did not end up receiving the mNutrition messaging, then the LATE estimates—which measure the average impact of receipt of the mNutrition messaging for complier households—may be a more policy-relevant reflection of the benefits of the service.

The LATE estimates use an indicator for whether the household ever received automatic nutrition text messages during the two years preceding the endline survey as the measure of service receipt.[5] To do so, they use the randomly assigned mNutrition treatment indicator as an exogenous instrument for whether the household received automatic nutrition messages on a mobile phone during the study period. Under the two assumptions discussed in the quantitative component report, two-stage least squares estimates that use the random mNutrition offer as the excluded instrument identify the causal impact of having received the mNutrition content for complier households.

4.3.5 Child anthropometry

As with the ITT estimates, the quantitative component was not able to reject the null hypothesis of no effect of the mNutrition service on child anthropometry. The point estimates are simply the ratio of the ITT point estimate to the first-stage coefficient, implying that they are between 2 and 3 times the size of the ITT treatment effects.

4.3.6 IYCF practices

LATE estimates of the impact of receipt of mNutrition messaging on IYCF practices for compliers remain statistically significant for the same set of IYCF practices as in the corresponding ITT table: continued breastfeeding at one year, children's dietary diversity for children aged 6–35 months, the

[5] The quantitative component used this measure instead of the self-reported information on receipt of the mNutrition messaging because of concerns that households were not accurately able to identify the sender of nutrition messaging. The treatment and control group differences in the likelihood of receiving messages from an unknown sender are evidence of this problem, as is the large gap between the early administrative data and the self-reported data. We therefore use the broadest possible self-reported information on message receipt as our measure of programme exposure.

likelihood of achieving MDD for children aged 6–35 months, and the likelihood that children aged 6–23 months meet minimum dietary adequacy.

The point estimates now indicate that for compliers, receiving the mNutrition messaging increased the likelihood that children aged 12–15 months were breastfed by 40.4 percentage points (standard error 0.162), or nearly 50% of the control group mean, though this result is based on a small sample and non-compliance in just two control households. For children's dietary diversity, receiving the mNutrition messaging increased the number of food categories that children consumed by 0.263 (standard error 0.148), or 7% of the control group mean. The MDD estimate for children aged 6–35 months now suggests that children in complier households were 9.5 percentage points (standard error 0.056) more likely to satisfy the MDD standard (15.7% of the control group mean). As with the ITT estimates, the improved dietary diversity scores seem to be driven by modest increases in the likelihood of consuming vitamin A-rich fruits and vegetables, other fruits and vegetables, and dairy products.

Receipt of the mNutrition service increased the likelihood that children aged 6–23 months in complier households had an adequate diet (MAD) by 17.9 percentage points in the basic model (standard error 0.088 p-value 0.043) and 21.0 percentage points in the extended model (standard error 0.090 p-value 0.019). These are impacts equivalent to 80.0 and 93.8% of the control group mean, suggesting the mNutrition service nearly doubled the number of children aged 6–23 months who received an adequate diet.

There was no statistically significant impact of mNutrition for compliers for either of the two IYCF practices Anderson indexes: both point estimates are positive, indicating a 0.021 increase in the index for children aged 6–23 months (p-value 0.251) and a 0.008 increase in the index for children aged 6–35 months (p-value 0.216).

4.3.7 Dietary diversity

The quantitative component estimates the LATE for compliers for WDDS, MDD-W, and HDDS. As was the case for the other LATE tables, the LATE estimates are between two and three times as large as the ITT estimates, indicating that the treatment effect for households that were induced to receive the content by the offer of access to the service are larger than the average treatment effects for all households that received the treatment offer. However, for WDDS the estimates are also slightly less precise, with p-values of 0.356 in the basic model and 0.319 in the extended model.

Similarly, the MDD-W results, though large in magnitude, are at best marginally statistically significant, depending on whether they use the basic or extended model. The basic model point estimate indicates that receipt of the mNutrition messages increased the likelihood that women achieved MDD-W by 8.8 percentage points (standard error 0.053 p-value 0.098), while the extended model suggests it increased MDD-W by 7.9 percentage points (standard error 0.051 p-value 0.125). Both point estimates are large relative to the control group mean, 14.1% and 12.6%, respectively.

The LATE estimates for HDDS, though positively signed and larger than the ITT estimates, are still small—at most, 1.9% of the control household mean—and have p-values of 0.870 and 0.641. **This further emphasises what was previously pointed out while discussing Table 4: the improvements in children's and women's diets were not driven by broader changes in household diets. Instead, the results are more likely due to changes in the way food was allocated across household members.**

4.4 LATE summary

The results suggest that receipt of the automatic nutrition messages during the previous two years had a larger impact on dietary diversity, IYCF practices, and knowledge of nutrition and IYCF practices for complier households than the offer of access to the mNutrition messages had for treatment households more broadly. With the offer of access to the mNutrition service increasing reported participation in the service by 35–40%, the LATE for complier households are between 2.5 and 2.85 times as large as the ITT estimates.

There are three principal reasons that could explain the gap in treatment effect sizes between the ITT and LATE estimates. First, self-reported receipt of the mNutrition content may have more of an impact on knowledge and behaviour than the offer of access to the content. Certainly, the authors of the quantitative component should not expect households that never received, noticed, or read mNutrition messages, despite being registered for the service, to update their beliefs or behaviour in any way. It immediately follows that they should therefore expect confirmed receipt of the content to have a larger impact than receiving an offer of access to the content for the same set of households.

In addition to comparing the impact of self-reported receipt to the impact of an offer to receive the content, the LATE estimates are also pertinent for a different set of households than the ITT estimates: namely, the LATE parameters measure impact for complier households, while the ITT estimates represent impacts for all households. With the indicator for ever having received automatic nutrition messages as the measure of receipt of the mNutrition content, the authors of the quantitative component know that complier households include roughly 40% of the full study population. If these complier households are also households that have higher expected benefits from the information in the mNutrition messages, then differences between the ITT and LATE estimates could be driven by the fact that they are average effects for different populations with distinct returns to the service.

Related to the two reasons mentioned above, it is important to note that the quantitative component measure of participation in the service—whether households self-report having received automatic nutrition-related text messages during the study period—is a noisy measure of true receipt of the mNutrition content. Some individuals and households will misremember or incorrectly report their true participation in the service, and this has implications for both the size of the complier population and the magnitude of the LATE estimates for complier households. If, for example, no control households received the mNutrition content and the administrative data on service participation in November of 2017 completely identifies the treatment households that received messages, then the LATE estimates would be 1.16 times as large as the ITT estimates, as opposed to 2.5–2.85 times the ITT estimates as presented. As the gap in mNutrition message receipt approaches one, the difference between the ITT and the LATE messages shrinks to zero as the complier population of the programme becomes the full study population.

In comparing the LATE estimates based on self-reports of receiving automatic nutrition messages during the study period, an important question is whether the LATE estimates or the ITT estimates are more relevant for policymakers when evaluating the usefulness and cost-effectiveness of the mNutrition service. Though the LATE estimates provide critical information about the potential benefits of the information contained within the mNutrition content, it is important, when calculating the benefits of the service, to consider that service take-up is imperfect even among households and individuals who express a willingness to participate. Whether imperfect take-up for treatment households is due to issues related to message salience, low attachment to SIM cards, the sharing of mobile phones or SIM cards across household members or even households, or network connectivity problems in remote, rural areas, these potential barriers to disseminating information

through mobile phones should be incorporated when comparing the effectiveness of mobile phone-based interventions to other methods of behaviour change communication. The quantitative component therefore preferred the ITT estimates to the LATE estimates shown here for characterising the service benefits.

4.5 User overview

The quantitative component focused on 2,477 fully complete interviews and a further 118 that were partially completed or had a missing male interview in the endline in the Iringa region. Key to a broader cost-effectiveness model will be the estimated profile of **user numbers**. This will lay the groundwork for estimating the impact on the whole of Tanzania, not just the study area that was the focus of the quantitative surveys. This section draws on the work of the business modelling component (Scott *et al.*, 2019).

The Tanzania mHealth-PPP submitted Wazazi Nipendeni registration data from the Rasello platform to the study team (authors). The data query was performed on 27 March 2019. The dataset contains the following fields:

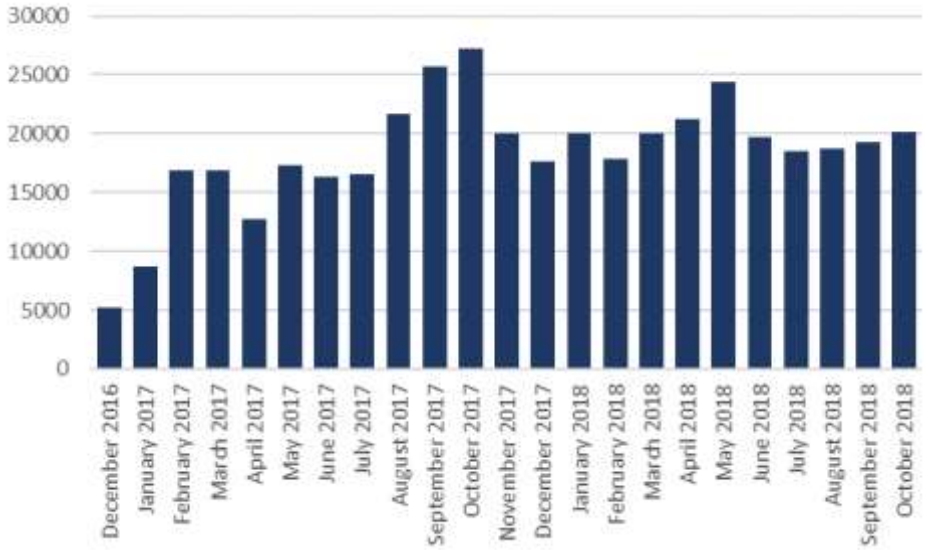
- unique ID (anonymised);
- source (means of registration, e.g. short code, API);
- partner (who registered the user);
- category of user, e.g. pregnant woman, mother with child;
- stage of pregnancy at registration;
- registration date;
- status (active or inactive); and
- opt-out date (if inactive).

The data required considerable manipulation, primarily because when pregnant women reach term (and are assumed to give birth), the original record is flagged as inactive and a new record is created, with a new category of mother with child. Given that most users are women who have given birth while registered as users, there were a large number of ‘paired’ records that needed to be reconciled and collated. The working dataset includes additional categories to reflect these hybrid classifications, e.g. PW=>MC represents a woman who registered when pregnant but has since given birth and has continued using the service as a mother with child.

4.6 Growth in user numbers

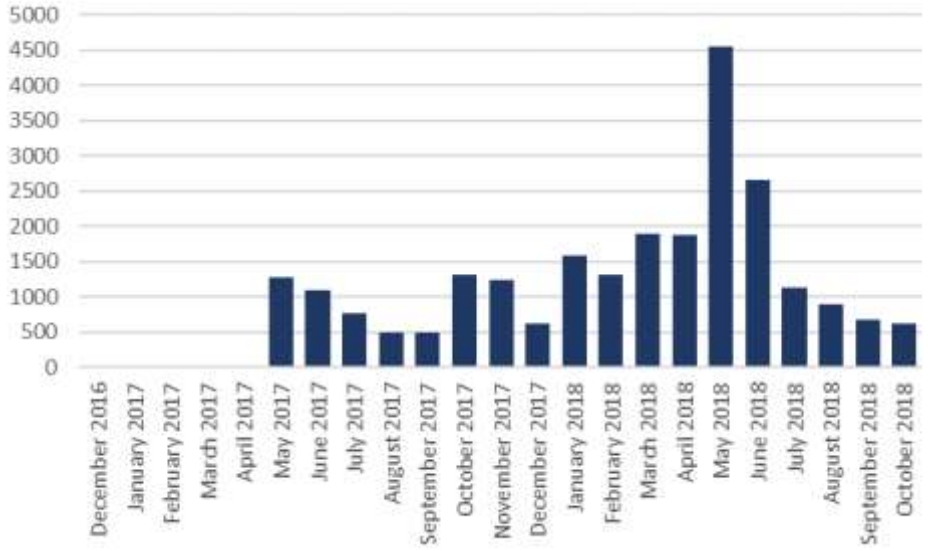
New registrations fluctuate monthly, as shown in Figure 3, and opt-outs fluctuate, as shown in Figure 4. In the period May 2017–October 2018 new registrations were running at an average of around 20,000 per month, whereas people were leaving the service at a much lower rate of less than 1,400 per month, i.e. 7% of new users. Combining new users with people leaving the service gives the cumulative profile of active users presented in Figure 5.

Figure 3: Monthly registrations



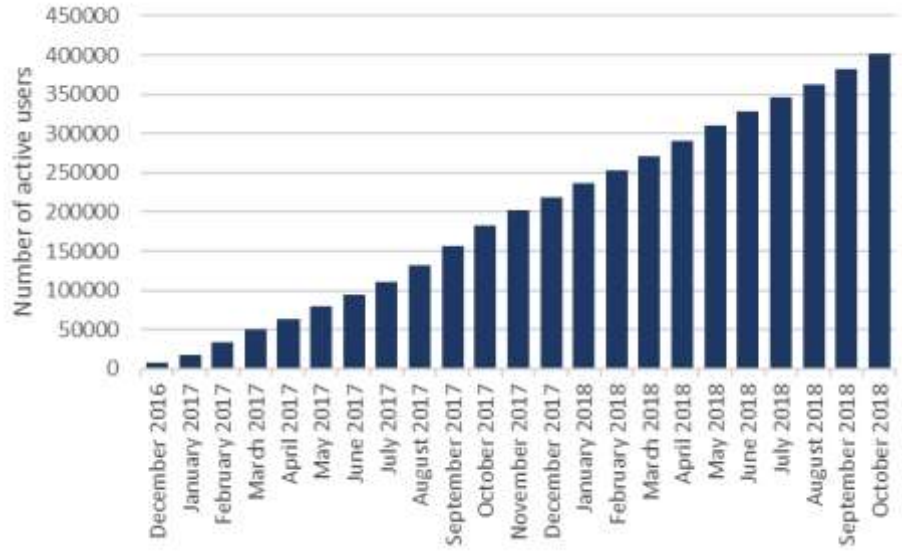
Source: Authors' own.

Figure 4: Monthly opt-outs



Source: Authors' own.

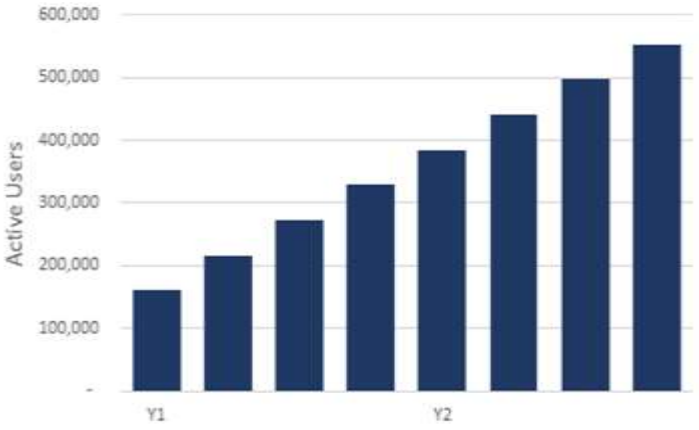
Figure 5: Active users



Source: Authors' own.

In the absence of reliable data, it is not possible to track patterns of growth in users from the introduction of the service. The total number of users may be up to 130,000 higher than these figures indicate, because many of the original users that were migrated to the new system have not been included in the above numbers.

Figure 6: Active users taking into account those migrated in December 2016



Source: Authors' own.

This was confirmed by the mHealth Tanzania-PPP, which estimated the number of active users to be 350,000 in November 2017. Adding 130,000 to the November 2017 estimate of 200,000 from Figure 6 gives a total of 330,000, which is consistent. Nevertheless, the patterns of growth and opt-outs remain accurate.

At October 2019, the date of this cost-effectiveness report, the number of current users was estimated as 550,000.

5 Effectiveness

5.1 Impacts identified by quantitative study

IYCF practices (ITT estimates)

Summarising Section 4, the quantitative component found for ITT estimates:

- a positive impact on the likelihood of continued breastfeeding for children aged 12–15 months, though there are only 27 children in the sample within this age range;
- a positive impact on **dietary diversity for children aged 6–35 months –** the service increased the number of food categories consumed by 0.107 (standard error 0.060 p-value 0.073) in the basic model;
- the mNutrition service increased the likelihood that children aged 6–35 months met the **MDD** – they were 3.8 percentage points more likely to have consumed from four food groups (standard error 0.023 p-value 0.093); and
- children aged 6–23 months were 6.9 percentage points more likely to satisfy **MAD** based on the model with basic controls (standard error 0.034).

Women's dietary diversity (ITT estimates)

Summarising Section 4, the quantitative component found for ITT estimates:

- the mNutrition service increased the likelihood that the primary females satisfied **MDD-W** by 4.0 percentage points in the basic model (standard error 0.021 p-value 0.062) – the effect sizes for MDD-W represent increases of 6.4% of the control group mean.

IYCF practices (LATE estimates)

Summarising Section 4, the quantitative component found for LATE estimates as follows:

- continued breastfeeding at one year: The mNutrition service increased the likelihood that children aged 12–15 months were breastfed by 40.4 percentage points (standard error 0.162), or nearly 50% of the control group mean;
- children's dietary diversity for children aged 6–35 months: The mNutrition service increased the number of food categories that children consumed by 0.263 (standard error 0.148), or 7% of the control group mean;
- the likelihood of achieving minimum dietary diversity for children aged 6–35 months: children in complier households were 9.5 percentage points (standard error 0.056) more likely to satisfy the MDD standard (15.7% of the control group mean); and
- the likelihood that children aged 6–23 months met minimum dietary adequacy: The mNutrition service increased the likelihood that children aged 6–23 months in complier households had an adequate diet (MAD) by 17.9 percentage points in the basic model (standard error 0.088).

Women's dietary diversity (LATE estimates)

Summarising Section 4, the quantitative component found for LATE estimates:

- the mNutrition service increased the likelihood that women achieved **MDD-W** by 8.8 percentage points (standard error 0.053 p-value 0.098).

5.2 Measurable change

In this section we consider how these changes translate into Disability Adjusted Life Years (DALYs).

In accordance with guidance given in the endline quantitative report, we focus on those impacts identified in the ITT analysis (**Assumption 1**).

The baseline design of the CEA suggested that effectiveness would be calculated based on one or more of the following:

- Anthropometric change: Across all the child anthropometry outcomes, the quantitative component found no evidence that providing households access to the mNutrition service had an impact—positive or negative—on any of the outcomes.
- WDDS: There is evidence that mNutrition increased dietary quality for both women of reproductive age and children, and that these increases were focused on individuals at or below thresholds that have been linked to dietary adequacy for these groups (MDD and MDD-W).
- Children's dietary diversity (IYCF practices): Three IYCF indicators show a significant increase, all of which relate to dietary diversity.

Regarding women's dietary diversity, in the baseline report the mechanism for 'translating' WDD to DALYs was said to be anaemia. In the Global Burden of Disease (see below), anaemia is a risk factor, and can be related to deaths and DALYs. At the time of the baseline report, there were a number of studies that suggested dietary diversity in women could be linked to anaemia. However, in the interim years, the link between anaemia and women's dietary diversity has been questioned (Saaka *et al.*, 2017), and the 'translation' of dietary diversity to DALYs is not considered straightforward.

Regarding children's dietary diversity, three IYCF indicators show a significant increase, all of which relate to dietary diversity. Any one of these could be used as a basis for estimating benefits due to improved dietary diversity, but not all, since using multiples of these indicators would lead to double counting. Secondary outcomes have not been included in the analysis as there is not a well-established link between knowledge and benefits to nutrition (to be calculated in DALYs – Leroy *et al*., 2018).[6]

5.2.1 Dietary diversity and stunting (children aged 6–35 months)

The CEA has, therefore, been based on estimating the benefit to be derived from an increase in **dietary diversity for children aged 6–35 months** of 0.11 food groups (**Assumption 2**).

Linking dietary diversity to stunting

Arimond and Ruel (2004) show that in four Africa countries (Ethiopia, Mali, Rwanda, and Zimbabwe), the number of food groups consumed had an effect on stunting (in children aged 6–23 months). They divided the number of food groups consumed into terciles:

- 0–2 food groups (midpoint = 1);

[6] 'DALYs can be used for health-related outcomes, but not for outcomes such as nutrition knowledge' (Leroy *et al*., 2018).

- 3–4 food groups (midpoint = 3.5); and
- 5–7 food groups (midpoint = 6).

The regression coefficients are summarised in Table 5.

Table 5: Summary of regression coefficients found in Arimond and Ruel (2004)

	Ethiopia	Mali	Rwanda	Zimbabwe
Coefficient (high to low)	0.35	0.23	0.17	0.68

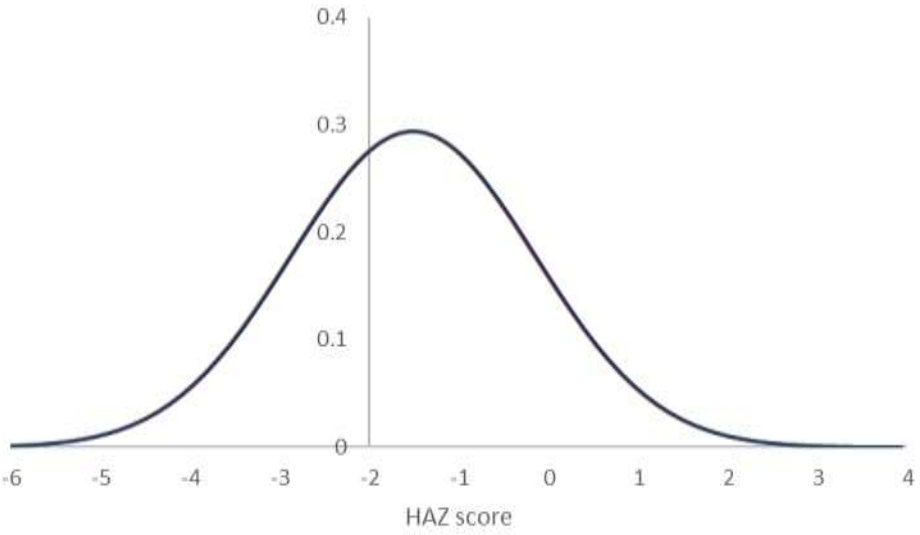
Source: Arimond and Ruel, (2004)

For the initial calculation of DALYs averted we take an average of these coefficients (0.36) (**Assumption 3**). We can thereby calculate that an increase of one food group will lead to a reduction in the HAZ score of 0.36/(6-1[7]) = 0.072 (**Calculation 1**).

The effect of increasing dietary diversity by 0.11 food groups corresponds to an effect on stunting of an increase in HAZ by 0.008 (**Calculation 2**).

A child is classified as stunted if their HAZ score is less than -2, i.e. their height-for-age ratio is more than 2 standard deviations below the WHO Child Growth Standards median. The weighted mean HAZ score for all children in Tanzania aged 6–35 months is -1.52, based on the DHS sample of 5,027 children in this age range (DHS, 2016). A standard deviation of 1.4 has been used for these HAZ scores (TFNC, 2014). This gives the distribution presented in Figure 7.

Figure 7: Distribution of HAZ scores for children aged 6–35 months



Source: Authors' own. Based on DHS data 2015[8]

The proportion of children lying below the -2 HAZ threshold in Figure 7 is 36.33%.

If the HAZ score for all children were increased by 0.008 the difference would be too small to discern on a distribution curve, but the curve would be shifted slightly to the right. In this instance,

[7] Difference in number of foods between low and high terciles.
[8] For the full data, see MoHCDGEC [Tanzania Mainland], Ministry of Health (MoH) [Zanzibar], National Bureau of Statistics, Office of the Chief Government Statistician, and ICF (2016).

the proportion of children lying below the -2 HAZ threshold would be reduced from 36.33% to 36.11% (**Calculation 3**).

Of the total population of Tanzania, 9.6% are aged 0–35 months. From the figures in Table 6, it is estimated that 7.9% of the total population are aged 6–35 months. In 2018 the population of Tanzania was 54.2 million (National Bureau of Statistics, 2019), so it is estimated that 4.28 million children are aged 6–35 months (**Assumption 4**).

Table 6: Tanzania Population figures (2012 census)

Age (years)	Number	% of total population (N = 44,928,923)
0	1,499,389	3.34%
1	1,349,091	3.00%
2	1,477,998	3.29%
Total	**4,326,478**	9.63%

Source: National Bureau of Statistics, 2013

We acknowledge that trying to link dietary diversity or nutrient adequacy to DALYs could be challenged, particularly in the light of the ITT and LATE that indicated there was no evidence that access to the mNutrition service had any impact on child nutrition. However, in doing so we make an assumption: that over the longer term the changes in diet diversity may cause changes in stunting (**Assumption 5**). The resulting analysis presented here is at risk of being generous to the programme and policy actors need to be aware of that.

Vitamin A deficiency

The quantitative component was not designed to assess levels of vitamin A deficiency, but it was able to show that children in the treatment communities were more likely to have eaten vitamin A-rich fruits or vegetables and other fruits. However, the Global Burden of Disease data do not attribute any burden of DALYs to children being at risk of a diet that is low in fruits, so it is not possible to enumerate any benefit in relation to this specific aspect of diet.

5.3 Calculating DALYs

The Global Burden of Disease data (Institute for Health Metrics and Evaluation, 2018a) are used as the reference for the burden imposed by various ‘causes’, where a cause is defined as ‘A single disease or injury or an aggregation of diseases and injuries that causes death or disability’ (Institute for Health Metrics and Evaluation, 2018b). The burden of a disease comprises years of life lost (YLL) and years lived with a disability (YLD), which are summed to give DALYs. For example, the total burden of disease shouldered by children under five years of age in Tanzania is 10.7 million DALYs.

A risk factor is defined as ‘An attribute, behaviour, exposure, or other factor which is causally associated with an increased (or decreased) probability of a disease or injury.’ The burden of a disease is not due to a risk factor, it is due to the disease, but a certain part of the burden can be attributed to a risk factor. For example, stunting is not classified as a disease, so there is no directly associated burden (DALYs). However, stunting is classified as a risk factor, meaning that stunted children are more likely to suffer from a range of diseases, each of which does have an associated burden. For each disease (or cause), the Global Burden of Disease data can attribute a proportion of the associated burden (DALYs) to associated risk factors. Not all of the burden of a

given disease can be attributed to risk factors. For example, only 0.7% of upper respiratory infections are associated with risk factors, whereas 91% of lower respiratory infections can be attributed to risk factors.

The health burden associated with the risk of being stunted for all children aged one month to four years is 295,500 DALYs (see Table 7). These figures are drawn from the Global Health Data Exchange for two age categories: post-neonatal (28–364 days), and one to four years. Dividing the burden in DALYs by the months covered by the age bracket indicates that the health burden is much higher among children in the lower age group (see Table 8).

Table 7: Tanzania Health burden : stunting (one month to four years) by cause

Disease (cause)	DALYs
Other infectious diseases	18,598
Enteric infections	55,855
Respiratory infections and tuberculosis	221,081
Total	**295,534**

Source: Authors' own

Table 8: Tanzania Health burden : stunting (one month to four years) DALYs

Age group	Post-neonatal (2–11 months)	One to four years
Burden from risk of stunting (DALYs)	140,082	155,452
Months covered by age group	11	48
Specific burden (DALYs/month)	12,735	3,239

Source: Authors' own

Using these specific burden figures, the burden on children aged 6 to 35 months is estimated to be 154,100 DALYs. (**Assumption 6**).[9]

A reduction in the rate of stunting from 36.33% to 36.11% across the nationwide population of children aged 6–35 months would reduce the number of cases of stunting by 9,440 (**Calculation 4**).

36.33% of the 4.28 million children in the age group is 1,555,058, i.e. 1.55 million children are affected by stunting. The reduction of these cases by 9,440 represents 0.61% of the baseline cases of stunting in this age group. If this saving were applied to the burden of disease associated with a risk of stunting among children aged 6–35 months it would reduce the burden of disease by 935 DALYs (0.61% of 154100 DALYs) (**Assumption 7 and Calculation 5**).

As of the end of 2018, there were roughly 550,000 active users on the Wazazi Nipendeni system. Even though 9% of active users were pregnant women, as opposed to mothers, it is assumed that all households will eventually have at least one child that will potentially benefit from an improvement in dietary diversity. It has therefore been assumed that the subset of the population that the intervention reached was roughly 550,000 children (**Assumption 8**). This represents roughly 12.8% of the 4.28 million children estimated to be aged 6–35 months.

Assuming that the composition of the health burden on children reached by Wazazi Nipendeni is similar to that on children nationwide, then the reduction in health burden can be calculated on a

[9] Assumption 6 assumes that 6 months 50% of the specific burden of DALYs associated with the group 2–11 months is assignable to the group 6–35 months, along with 24 months of the specific burden of the one to four years age group.

pro rata basis (**Assumption 9**). If this assumption were true, then Wazazi Nipendeni would have been estimated to have reduced the health burden by roughly 119 DALYs (12.8% of 935 DALYs).

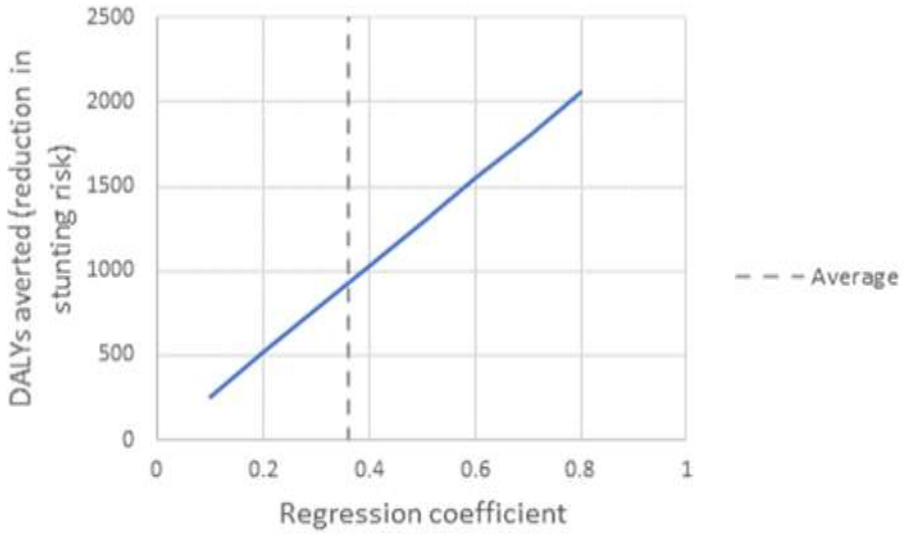
However, Assumption 2 can be challenged. The ITT considered the difference between the treatment group as a whole and the control group. Within the treatment group the take-up is 66.4%. When we consider the 550,000 users these are users who have taken up the service. So, while the ITT reports a change in dietary diversity for children aged 6–35 months of 0.11 food groups, this is for the whole sample – those who have taken up the service and those who have not. Similarly, in the control group 26.7% stated that they received health messaging. For this reason, the LATE estimations are closer to representing the 550,000 users and we choose to use the LATE estimations on the dietary diversity changes in the age group. The LATE estimation calculated by the quantitative component increases the number of food categories that children consume by 0.263 (standard error 0.148). This is more likely to represent the change found in the 550,000 users (**Assumption 10**).

Given this focus on the treatment take-up, the DALYs averted rises to an upper bound of 286 DALYs (**Calculation 5**).

5.4 Sensitivity analysis

The averaging of the findings of Arimond and Ruel could be interpreted differently. If we take the extremes of Rwanda and Zimbabwe, how sensitive are the total DALYs numbers to the correlation coefficient? Figure 8 shows that the number of DALYs could potentially halve or double.

Figure 8: Sensitivity of the total DALYs numbers to the correlation coefficient



Source: Authors' own.

6 Cost breakdown

This section presents the costs collected that will be used for Analyses A and B. The section undertakes a sensitivity analysis on the key costs. These costs have relevance for the business modelling component and can be found in the business modelling report (Scott *et al*., 2019).

6.1 Costs for Analysis A

Costs were collected from project budgets, expenditure reports, and key stakeholder contacts from multiple organisations as available from inception to July 2019. If a PPP in another country were to consider adding nutritional messaging to an existing health platform, they would need to consider the costs associated with establishing a localised information database, along with the capital expenditure and operational expenditure required to get such a system up and running.

Where possible, the team has collected and collated costs for the Wazazi Nipendeni service. The service is being treated as a whole, therefore any effect is the combination of, or synergy between, health messages and the nutritional messages. We have therefore at this stage taken the costs for the whole.

In the framework, the setup and ongoing costs were said to include:

- **Capital costs**: The cost of any infrastructure created to support Wazazi Nipendeni. In order to provide the service some extra contracted aggregation was required. For instance, Rasello was contracted as the platform of delivery. A service offering in a new country might require some capital equipment but this is assumed to be covered within the aggregated contract.
- **Management/personnel costs**: The ongoing service requires expenditure on staff and management. PPP overheads could be incorporated here. Personnel costs need to include any data analysts required to maintain the platform. In each case, the staff costs stated in the budget and reporting documents are attributed to the associated activity.
- **Promotion and marketing**: This includes the training of in-country personnel, transport for trainers, hours of labour etc. Wazazi Nipendeni has already taken a number of different approaches to sharing, making alliances with NGOs throughout the country who are undertaking health campaigns.
- **Recurrent costs of messaging**: On the face of it, one of the simplest costs is the price assigned to the text messaging. Each message has a cost associated with it. Message scheduling and dispatch platforms will also incur ongoing maintenance costs. Who pays these costs is a more complex question. For the duration of the study, these costs were covered by the MNOs. Please note that while the PPP budget assigned the retail price of the SMS as a contribution from the MNO, and that is reflected in the tables below, the actual cost of an SMS is debatable and is discussed in Section 8.
- **Localisation content development**: mNutrition as a whole has been funded to develop and collate a global repository of nutrition information. In order for this to be applied to Wazazi Nipendeni there had to be a localisation process: taking the global fact sheets and making them relevant to the clientele of Wazazi Nipendeni.
- **Content curating**: There is an ongoing need to update the content of the messages. Information can get out of date and while this is more likely to happen with medical information, there is nevertheless a need to ensure that the health and nutritional information remains relevant.

- **User experience, baseline, and monitoring and evaluation**: Resources and personnel needed for user experience surveys and feedback (called 'UX' by the industry), baseline surveys, and monitoring and evaluation. We include here the baseline surveys and user experience surveys required to design the specifics of the service, and the ongoing mechanisms of feedback to keep the service relevant to the farmers and to keep stakeholders apprised of the services effects (both financial returns and public good impact). If a similar service utilising the experience of Wazazi Nipendeni and the global content created by the mNutrition were to be set up in another country, there would need to be further user experience surveys to inform the service shape and form, and to contribute to the localisation of the content.

6.1.1 Cost data utilised for the financial model

In the baseline report (Scott *et al*., 2018), it was proposed that the relationships between key stakeholders could be regarded as a multi-sided platform business model. This provides a means of making a product free to one group of customers, while another group pays. When considering financial viability, the principal cost components are as follows:

Fixed costs:

- developing content (Global Content Partnership and localising of content);
- mNutrition programme support (e.g. GSMA business intelligence); and
- developing the Wazazi Nipendeni service (Tanzania mHealth-PPP).

Operating costs:

- mHealth Tanzania-PPP (offices, staff, servers);
- government agencies (staff costs);
- SMS and USSD costs (MNOs); and
- field partner agencies (staff costs and expenses).

Table 9 and Table 10 present the core costs and the sources of information for each.

6.1.2 Capital expenditure

Table 9: Capital expenditure

Item	Estimate (£)	Source	Description
Country-level investment			
Localisation of content	62,500	CABI budget	All local content partner payments allocated to Tanzania (62,468)
Staff costs (global content partners)	20,000	CABI budget	All GAIN staff costs allocated to Tanzania (20,303)
Direct costs	0	CABI budget	All GAIN direct costs allocated to Tanzania (0)
mNutrition programme (Country-specific)			
Product development			
Formative evaluation (international)	100,000	GSMA communications	See average country breakdown below
User experience expert and design consultants	130,000	GSMA communications	See average country breakdown below

Source: Authors' own

6.1.3 Operational expenditure

The following operational estimates are based on data supplied by Cardno.

Table 10: Operational expenditure

Item	Estimate (£)	Unit	Source	Description
Fixed costs				
Product development				
Content curation (local content partner)	34,000	£/year (after launch)		Lump sum estimates for keeping content up to date
Marketing expenses				
Marketing events	0	£/qtr		No provision is made for direct marketing expenditures, as marketing is now done mostly by partner organisations through face-to-face contact and campaigns that are part of other field programme activities, i.e. cannot be allocated directly to Wazazi Nipendeni
Administration expenses				
Salaries and wages	65,000	£/qtr	Cardno communication	Assume 50% effort dedicated to Wazazi Nipendeni ($55,784/month)
Fringe benefits	60,000	£/qtr	Cardno communication	Assume 50% effort dedicated to Wazazi Nipendeni ($51,777/month)

Equipment	0	£/qtr	Cardno communication	0
Supplies	2,000	£/qtr	Cardno communication	Assume 50% expenses dedicated to Wazazi Nipendeni ($1,693/month)
Contractual (Rasello)	82,500	£/qtr	Cardno communication	Assume 50% expenses dedicated to Wazazi Nipendeni ($71,267/month)
Travel	1,750	£/qtr	Cardno communication	Assume 50% expenses dedicated to Wazazi Nipendeni $1,442/month)
Other	4,000	£/qtr	Cardno communication	Assume 50% expenses dedicated to Wazazi Nipendeni ($3,340/month)
NGO partners (training and development)	7,700	£/qtr	Cardno communication	Estimate of staff time ($3,333/month)
Variable costs (cost of sales)				
Cost of SMS (Vodacom)	57.6	TZS/SMS	mHealth Tanzania-PPP report	SMS market prices
Cost of SMS (Airtel)	69	TZS/SMS	mHealth Tanzania-PPP report	SMS market prices
Cost of SMS (Tigo)	79	TZS/SMS	mHealth Tanzania-PPP report	SMS market prices
Cost of SMS (Halotel)	33	TZS/SMS	Halotel Website	Prevailing market rate for individual customers
USSD sessions	120	TZS/session	mHealth Tanzania-PPP report	Single price attributed to all MNOs

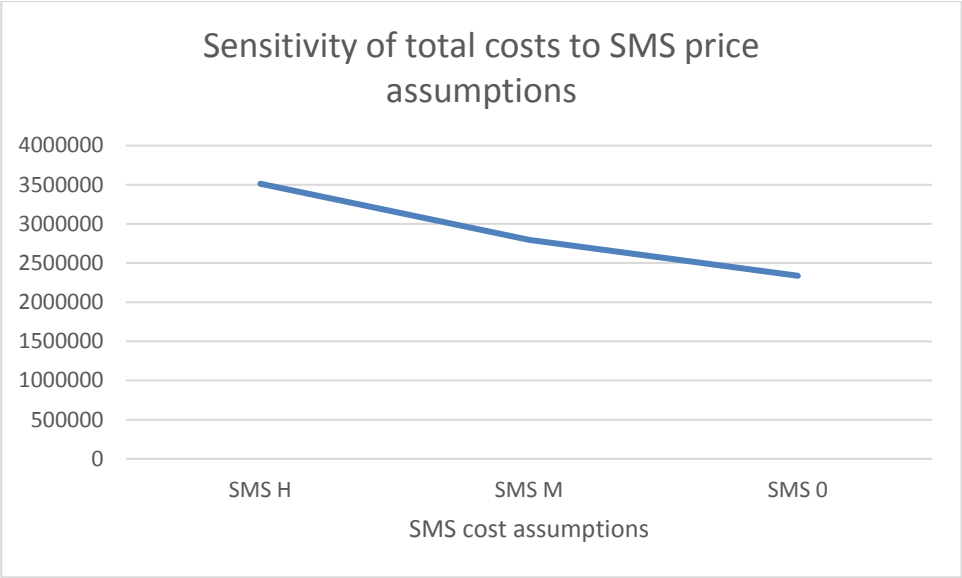
Source: Authors' own

6.1.4 Cost assumptions for the analysis

Cost assumption 1. The SMS are donated by the MNO, and the real cost to the MNO could be taken as anywhere between a nominal amount (0.003 US cents) and the retail price.

The total cost assigned to the programme (Analysis A) is sensitive to the costs assigned to the SMS (Figure 9).

Figure 9: Sensitivity of total costs to SMS price assumptions



SMS H – retail prices assigned to MNOs as reported by Cardno.

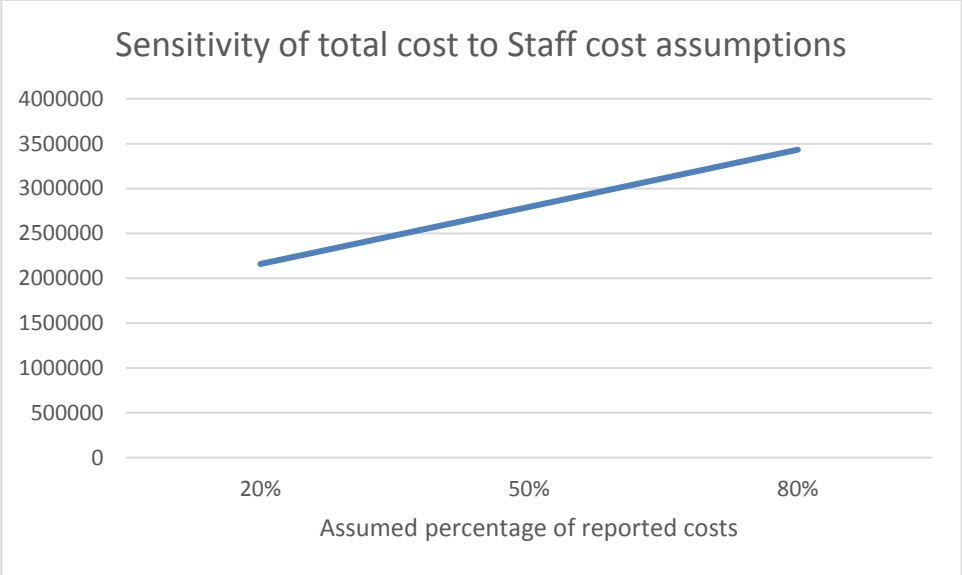
SMS M – an assumed bulk purchase price, (estimated at TZS 25 per SMS), being the same across all MNOs.

SMS 0 – a nominal assignment of costs to SMS, as accounted for within MNOs.

Source: Authors' own.

Cost assumption 2. The management costs can also be queried. The Cardno reports assigned salaries and wages and fringe benefits in full to Wazazi Nipendeni. However, Cardno manages much more than Wazazi Nipendeni, and even within Wazazi Nipendeni there are activities that go beyond the mNutrition message sharing. The team have assigned 50% of the costs of Cardno to the service on the assumption that if the service were replicated without the synergy of all that Cardno does, the staff would have to be in place for Wazazi Nipendeni to get the same or similar reach and to develop the partnerships required with NGOs, government, and MNOs. The total cost assigned to the programme (Analysis A) is sensitive to the costs assigned to the salaries and wages and fringe benefits (Figure 10).

Figure 10: Sensitivity of total cost to staff cost assumptions

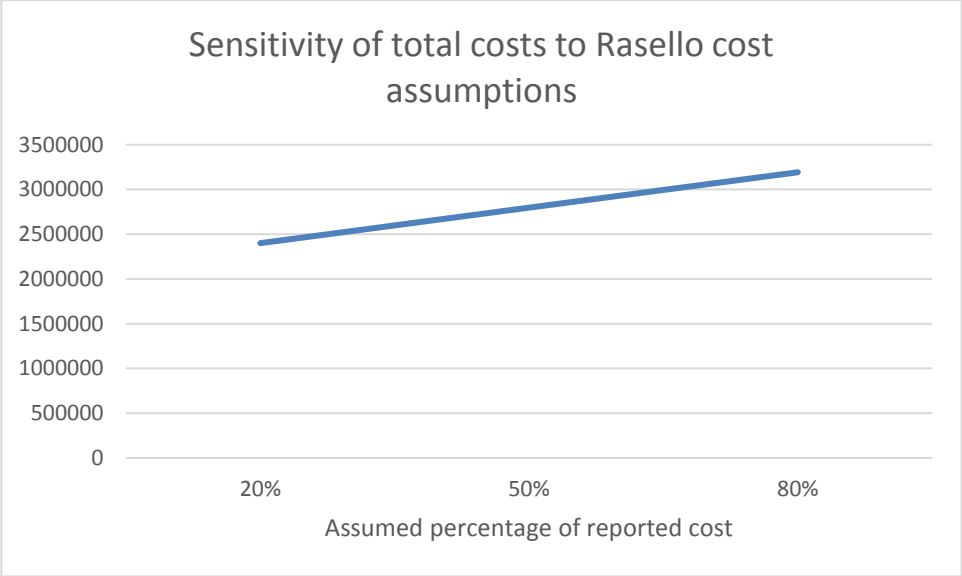


Source: Authors' own.

Cost assumption 3. Rasello also undertakes work other than the mNutrition messaging, and its contract reflects that. Again, the team have assumed 50% of Rasello's contract is assigned to Wazazi Nipendeni.

The total cost assigned to the programme (Analysis A) is sensitive to the costs assigned to the Rasello contract (Figure 11).

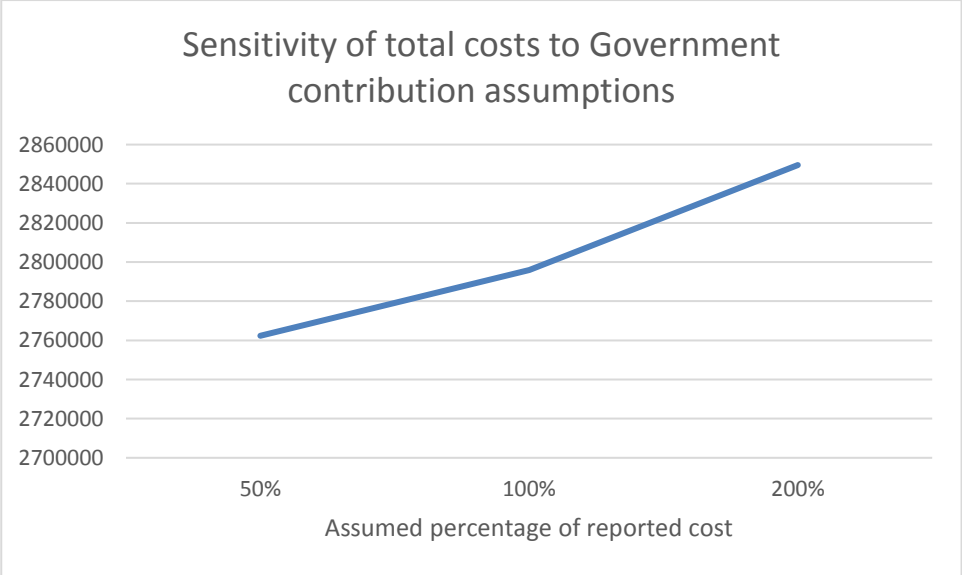
Figure 11: Sensitivity of total costs to Rasello cost assumptions



Source: Authors' own.

Cost assumption 4. The contribution of the Government of Tanzania is based on interviews and budget data supplied. Given that this is not drawn from a financial system per se, the costs could vary. Figure 12Figure 11 shows the sensitivity of the total to the assumption made about the Government staff cost contribution. Note the Y axis does not start at 0.

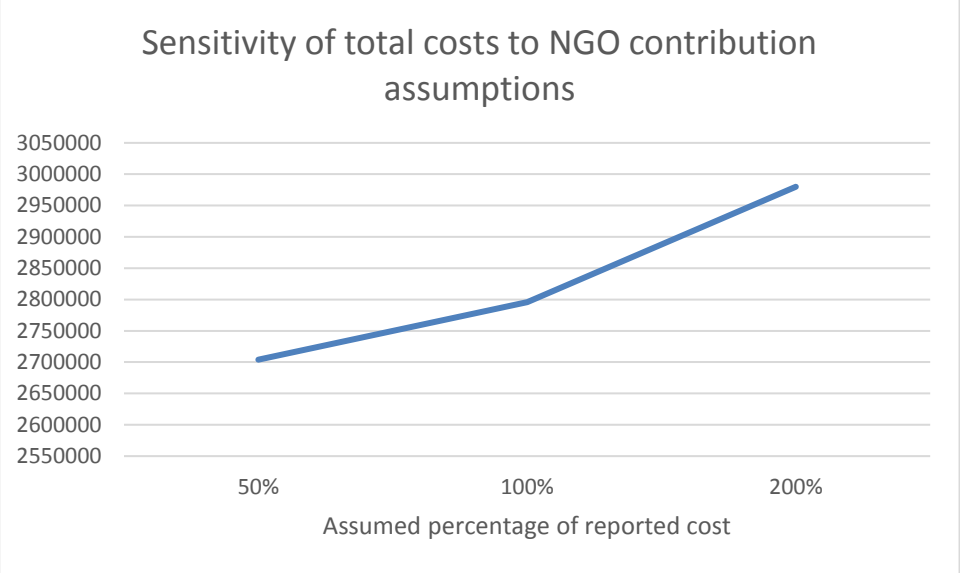
Figure 12: Sensitivity of total costs to Government contribution assumptions



Source: Authors' own.

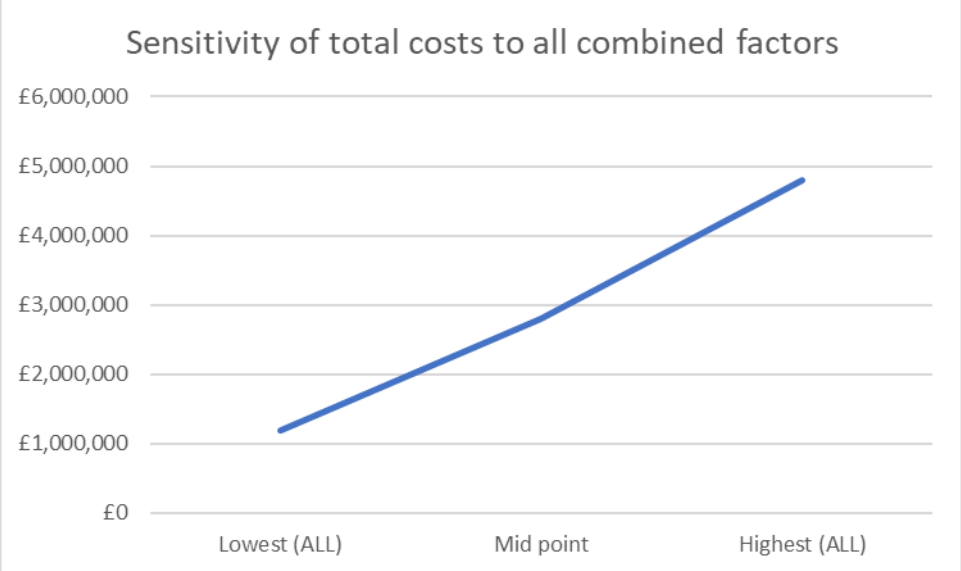
Cost assumption 5. The contribution of the NGO partners is based on interviews and budget data supplied. Given that this is not drawn from a financial system per se, the costs could vary. Figure 13 shows the sensitivity of the total to the assumption made about the NGO partner cost contribution. Note the Y axis does not start at 0.

Figure 13: Sensitivity of total costs to NGO contribution assumptions



Source: Authors' own.

Figure 14: Sensitivity of total costs to all combined factors



Source: Authors' own.

The total cost can therefore be taken as between £1,182,000 and £4,792,000 (Figure 14).

6.2 Analysis A

Accepting the Assumptions 2 and 3 at 50% and assigning a bulk purchase price of 25 TZ Shillings to the price of SMS, the cost of the programme for Analysis A will be taken at £2,796,000 for the two years.

For the 550,000 current users (two years of added nutritional content), this suggests the current costs are £5.08 per user (a range of between £2.14 and £8.71 per user).

The ITT estimates 119 DALYs averted, while the LATE estimates 286 DALYs averted. Taking the LATE estimates as more representative of the effect on users, 286 DALYs suggests a cost per DALYs averted of £9,776 (a range of £4,132 to £16,755 (2019 prices)). As a sensitivity analysis, the ITT estimates would present a mean cost per DALYs averted of £23,495 (a range of £9930 to £40,268).

Even at the lowest estimate, this exceeds the WHO guidelines of a cost-effective intervention (three times GDP; the GDP of Tanzania is £622 per capita).

6.2.1 Time horizons for Analysis A

The above analysis has been undertaken for the two-year period of study.

If we consider a four-year horizon, the total cost would be £5,886,000, reaching just over 1 million users (assuming a similar growth to the study period), and if we make the assumption that there is an equivalent increase in DALYs in the coming two years (572 LATE), the cost per DALYs averted would be £10,290.

If we consider a six-year horizon, the total cost would be £9,553,000, reaching just over 1.45 million users (assuming a similar growth to the study period) and if we make the assumption that there is an equivalent increase in DALYs in the coming two years (858 LATE), the cost per DALYs averted would be £11,134.

6.3 Costs for Analysis B

For the study to inform thinking on replicating the mHealth + Nutrition information service, actors need to be able to understand the contribution or investment made by FCDO in the wider mNutrition programme and its management by GSMA. This approach is helpful for donors and policy actors who are interested in assessing whether the programme represents value for money. In order to include this wider perspective, the following costs are explored, in addition to those costs included in Analysis A:

- **Research and development for the mNutrition programme as a whole:** The expanded nutritional messages sent out by Wazazi Nipendeni are only one particular output from the wider mNutrition programme. The mNutrition programme as a whole has spent time strategising, planning, co-creating global content, etc., leading to 14 specific in-country services in 12 countries. While it is impossible to extract the specific costs of mNutrition related to the new nutritional content of Wazazi Nipendeni, it could be argued that 1/14 of the overall programme costs (minus specific grants) should be imputed to the mNutrition component of Wazazi Nipendeni. This should capture the research and development behind the mNutrition project after the project's inception (hours of labour devoted to the project by larger organisations, the amount paid to external researchers, costs of rent, vehicles and other transport costs associated with the project, costs of office supplies, electricity and other expenses necessary for research and development tasks).
- **Global content development:** The mNutrition programme paid for a global content generation process that was carried out by a consortium, comprising CABI, GAIN, Oxfam, the International Livestock Research Institute, and the British Medical Journal. The global content partnership was responsible for identifying relevant content, creating content structures, and specifying

content validation and quality control processes. The content developed by the consortium was then made available to local content partners in each country to adapt for local consumption, and these costs are included as the localisation content development in Analysis A. A proportion of the costs associated with the work of the global content consortium could be imputed to Wazazi Nipendeni in Analysis B.

- **Project-related infrastructure:** In order to implement a complex programme across 14 projects in 12 countries and two continents, GSMA had to set up substantial management structures, at substantial cost. This includes institutional management structures, personnel, offices, IT networks etc. It is a proportion of these costs, paid for through the wider mNutrition programme, that are included in Analysis B.

6.3.1 Global content development

Table 11: Costs associated with the development of the wider mNutrition programme

mNutrition programme (global)			
Global content development			
Global content partners	250,000	CABI budget	255,910 (per country programme)[10]
Programme management			
Business intelligence and programme management (GSMA)	480,000	GSMA communications	See average country breakdown below

Source: Authors' own

GSMA mHealth total project budget average per country:[11] £1,055,900. However, in Analysis A we have already assigned £315,000 to the country localisation and user experience. The balance of £740,900 can be assigned to global content development; £255,900 (from the CABI budget for CABI costs) and £477,000 covers GSMA contributions to business intelligence and programme management.

The CABI budget outlines costs associated with the local content generation process, which includes both payments to local content partners and consortium staff costs (these are included in Analysis A, Table 9). However, additional costs are allocated to the consortium partners for direct costs and staff costs, which amount to over £3.5 million. A crude assumption can be made that these are spread evenly across all 14 projects, giving a total of £256,000 per country, or £128,000 per year if split over two years. The cost of content development for mNutrition was particularly high, as it was premised on building capacity within local institutions. It has been argued that it would have been possible to develop content at a cheaper rate had the capacity building mandate not been in place.

6.3.2 mNutrition programme as a whole

GSMA has provided an estimate of the average total budget per project of £1,055,900 for those countries running mHealth projects as part of mNutrition. We have identified direct expenditure items which are included in Analysis A (localisation content development, and product

[10] An alternative document suggests that one fifth of all HQ staff payments made to GAIN (£23,181) could be assigned to the Wazazi Nipendeni element of the global content production. However, if this is done the balance of the difference would have to be assigned to the Business Intelligence and Programme management. There is therefore no reason to use the figures from this second document.
[11] Personal communication.

development (monitoring and evaluation, user experience expert and design consultants, and business intelligence)), and the proportion of global content development. When these items are deducted from the average budget spend, the balance is £477,000. This has been split evenly over two years. This does not take into account whether GSMA has costed its overheads commercially. It is more than likely that some other parts of GSMA are subsidising the mDevelopment stream.

With these added costs for Analysis B, a range of £1,789,000 to £5,031,000 is estimated.

6.4 Analysis B

Accepting Assumptions 2 and 3 at 50% and assigning a bulk purchase price of 25 TZ Shillings to the price of SMS, the cost of the programme for analysis A will be taken as £3,526,000 for the two years.

For the 550,000 current users (two years of added nutritional content), this suggests the current costs are £6.41 per user (in the range of between £3.47 and £10.04 per user).

The ITT estimates 119 DALYs averted, while the LATE estimates 286 DALYs averted. Taking the LATE estimates as more representative of the effect on users, 286 DALYs suggests cost per DALYs averted of £12,328 (a range of £6,684 to £19,308 (2019 prices)).

Even at the lowest estimate, this exceeds the WHO guidelines of a cost-effective intervention (three times GDP; the GDP of Tanzania is £622 per capita).

6.5 Costs for Analysis C

To obtain a more complete picture of the costs involved in the Wazazi Nipendeni service, Analysis C, presented in the baseline framework, proposed considering a number of additional costs:

- indirect, variable costs incurred as a consequence of users taking actions to implement new practices advocated by the Wazazi Nipendeni service; and
- sunk costs involved in building the assets that each partner brings to the delivery of the Wazazi Nipendeni service.

The benefit or effectiveness of the Wazazi Nipendeni service was, in theory, potentially dependent on other costs to be met by the household. Such changes in health costs were not significant within the quantitative component – for example, the number of antenatal clinic visits by households in the ITT is not a significant increase over the control group.

Similarly, Analysis C was going to consider costs associated with implementing changed behaviour, as well as establishing intellectual and infrastructure assets that are employed in some way as part of achieving nutritional outcomes. Each of the invitations to utilise Wazazi Nipendeni were part of wider behaviour change communication campaigns across more than 23 government and NGO programmes. The study was not able to capture the background costs of each of these interventions and to assign a proportion to the CEA.

Finally, the quantitative research found that the Average Revenue Per User (ARPU) was TZS 510 higher among people who used Wazazi Nipendeni (compared with those who did not use the service). Given an ARPU of TZS 5,300 per month (£1.90/month) among the control sample, this indicates that the VAS stimulates a 10% increase in ARPU. It is not known whether the additional ARPU expenditure is spent on health-related activities as a consequence of the messaging or is a

function of increased familiarity with the phone. The qualitative research found that women became more comfortable and confident in using their phone when receiving messages from Wazazi Nipendeni, implying that it is the latter. The literature on phone usage among the poor in Sub Saharan Africa suggests the mobile penetration tends to be pro-poor and tends to have a positive effect on household finances. (e.g. Asongu 2015, Aker et al 2016, Danquah and Iddrisu 2018). This suggests that the net effect on household finance from any increased ARPU could be positive. Taking the increased ARPU expenditure as an indirect cost for the Wazazi Nipendeni would be an unreasonable assumption. Given the quantitative study did not identify any significant change in any single category of household expenditure, it would also not be reasonable to suggest that the increased ARPU is leading to decreased expenditure of food or to assume it is a burden to household finances. Given the uncertainties surrounding the source of the extra ARPU expenditure, and the consequences of its use and net benefit to the household, it is not possible to consider it an "*indirect, variable costs incurred as a consequence of users taking actions to implement new practices advocated by the Wazazi Nipendeni service*" (Analysis C, Baseline model)

Therefore, Analysis C was not attempted.

7 Commercial sustainability

The terms of reference for the impact evaluation include a request for a comment on the commercial sustainability of the intervention (because the study includes traditional MNO-operated VAS). In the case of Wazazi Nipendeni it is difficult to argue a case for the financial viability of the mHealth-PPP because none of the partners generates revenue directly from providing the Wazazi Nipendeni service. Indeed, all parties are prohibited from generating direct revenue because the Government of Tanzania has ruled that all health services should be free to consumers. However, the quantitative research has found that ARPU increased among users, so it is possible to develop a business case for MNOs to provide an mHealth service on the basis of indirect benefits should this be the model that is transferred to another country. If there is a commercial case for MNOs to provide an mHealth service, then there is also a case for a third party to provide content on a revenue share basis (as with conventional VAS).

Under the current business model, MNOs made no direct investment in developing the Wazazi Nipendeni service, but they donated the SMS messages and USSD sessions that the service depends on. Without any investment, standard financial metrics such as IRR or payback period are not relevant. In this case the contribution margin[12] gives an indication of financial performance, or the contribution that the product makes to group profit.

In the baseline report (Batchelor *et al*., 2018), it was proposed that the relationships between key stakeholders could be regarded as a multi-sided platform business model. This provides a means of making a product free to one group of customers, while another group pays. This analysis goes on to consider the financial viability of two scenarios:

- An independent but commercial content provider: The provider invests in localising content (assuming nutrition and maternal health content is in the public domain) and manages both the technical systems and arrangements with national health programmes (in the same way as the mHealth Tanzania-PPP). It provides information services as a VAS to MNOs on a revenue-sharing basis.
- An in-house MNO service: The MNO invests in localising content (assuming nutrition and maternal health content is in the public domain) and sets up an internal product development group to manage technical systems and to make alliances with national health programmes (in the same way as the mHealth Tanzania-PPP).

A financial model has been created to explore each of these options, based on cost data submitted by the mHealth Tanzania-PPP, Cardno, CABI, and GSMA.

7.1 The financial model

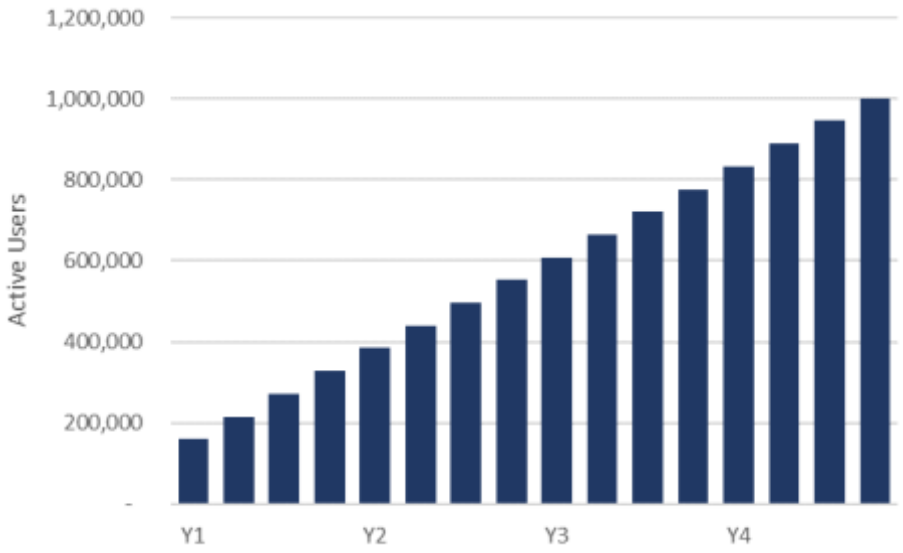
The model is based on operating cost structure and cost data provided by Cardno, supplemented by data gathered through interviews. These cost data include a quantification of in-kind contributions made by MNOs (donating SMS and USSD sessions) and partner NGOs (mainly time).

Key to the model is an estimated profile of user numbers, which is the basis of calculating revenues (from indirect benefits) and variable costs. The profile of active users in Figure 5 suggests numbers are continuing to rise, given roughly constant rates of new subscriptions, and

[12] Per unit revenue (indirect benefit) – variable costs. The contribution margin is often expressed as a percentage, when divided by the per unit revenue.

relatively low rates of opt-outs. This is likely to be the case for at least five years, at which time early users will start to leave the system as their children reach five years of age. It is estimated that there were over 100,000 active users on the system at the beginning of the period for which reliable data are available, which were used to calculate the trends in Section 5. The financial analysis is based on a four-year period, during which the number of active users can be expected to continue to rise as illustrated in Figure 15.[13]

Figure 15: Estimated user numbers profile for four-year period



Source: Authors' own.

Cash flow is calculated from estimates of revenues and variable costs, which depend on the customer numbers profile, as well as fixed costs.

Cash flow (operating profit) = revenue – cost of sales – fixed costs

Revenue:

- indirect benefits – increase in ARPU.

Cost of sales:

- cost of SMS – nominal value to MNO of text messages sent to customers (disaggregated by MNO); and
- cost of USSD sessions.

Fixed costs:

- administration expenses (mHealth Tanzania-PPP) – project management, staff costs, platform, travel;
- government institutions – in-kind staff costs;
- NGO partners – in-kind staff costs; and
- product development – content curation.

[13] Given the concerns over the quality of the early data, the number of active users at November 2016 has been rounded down from 130,000 to 100,000.

The analysis considers the viability of the service from various perspectives, including the MNOs under the current arrangements, but also possible commercial scenarios. Therefore, it considers investment costs incurred in developing the product for roll-out in a specific country. The following investment costs have been included in the model:

- global content development;
- localisation of content – e.g. developing appropriate messages, seeking necessary approvals;
- support service provided by the mNutrition programme – formative evaluation, user experience consultants.

Note that no direct grant payments were made to the mHealth Tanzania-PPP under the mNutrition programme.

7.2 Public good model

Under the current arrangement, the Tanzania mHealth-PPP is funded by public money (from international donors), and the service is enabled by the charitable donation of SMS messages by the MNOs. The costs of developing the Wazazi Nipendeni service have been covered by CDC, as part of its public health programmes. The costs of developing the additional nutrition content have been covered by the mNutrition programme. The service is funded and regarded as providing a public good, which it undeniably is. However, the quantitative study has shown that the MNOs are reaping tangible, if indirect, financial benefit from the service. This section considers the value of this benefit to partner MNOs under the current arrangement.

Financial reporting by the mHealth Tanzania-PPP quantifies the financial contribution made to Wazazi Nipendeni by each of the supporting MNOs. It does this by estimating the number of SMS messages donated by each MNO and then multiplying by the consumer retail price of SMS messages. In the absence of any payment to Wazazi Nipendeni for the content, the cost of messages is the only variable cost, but it varies between MNOs according to the different SMS retail prices.

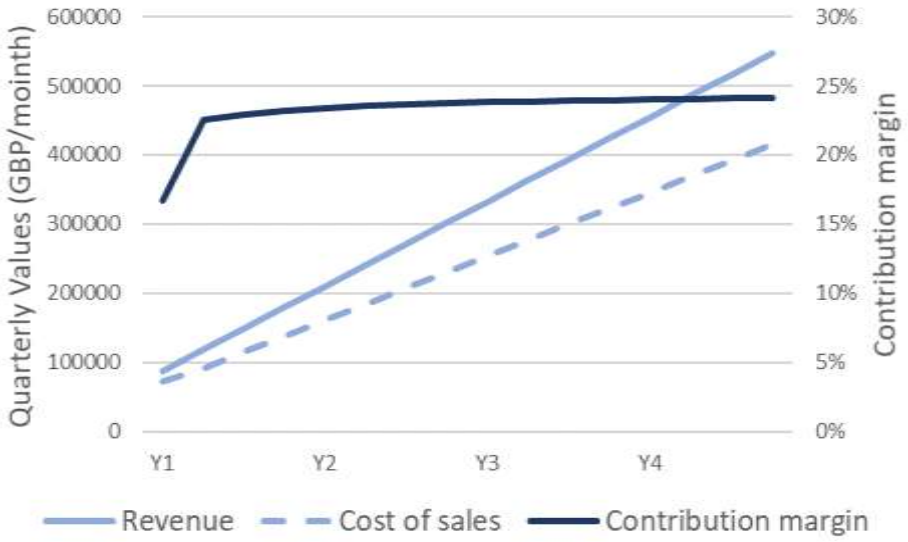
As stated above (Section 6.5) the quantitative research found that ARPU was TZS 510 higher among communities offered Wazazi Nipendeni (compared with control communities). Given an ARPU of TZS 5,300 per month (£1.90/month) among the control sample, this indicates that the VAS stimulates a 10% increase in ARPU. This is supported by findings from the qualitative research that found women became more comfortable and confident in using their phone when receiving messages from Wazazi Nipendeni, implying that they then used their phone more, using more airtime. We have made an assumption that the ARPU difference between users and non-users is due to Wazazi Nipendeni.

Given the user growth profile in Figure 15, the contribution margin flattens out at 24% (see Figure 16). The cost of SMS messages is directly proportional to the number of active users. However, the cost of USSD sessions relates to the number of new users, which is assumed to be constant each month, despite the fluctuations evident. At high user numbers, this cost component becomes a much smaller proportion of variable costs, dropping from 9% of variable costs in Year 1 quarter 1 to 1% of variable costs in Year 4 quarter 4. This explains why the contribution margin is lower at lower user numbers.

Note that the variable costs in Figure 16 are based on the aggregated costs across all four MNOs, each of which has been allocated different charge rates for SMS messages according to their retail

prices (at baseline). Table 12 shows the sensitivity of the contribution margin to the SMS price (at high user numbers).

Figure 16: Contribution margin



Source: Authors' own.

Table 12: Contribution margin for each operator (asymptotic)

MNO partner	SMS price (TZS/SMS)	Contribution margin (Year 4)
Vodacom	57.6	29%
Airtel	69	15%
Tigo	79	3%
Halotel[14]	33	61%

Source: Authors' own

The price attributed to sending SMS messages is clearly highly influential in determining how attractive their contribution to Wazazi Nipendeni appears to operators. So, what is the cost to an operator of sending an SMS? In a study on the impact of consumer consumption of network services (voice, SMS, and data) on network profitability, Blackburn *et al*., (2013) highlight a trend towards flat-rate pricing (as opposed to pay as you go). They point out that one feature that makes this attractive to operators is that the cost of delivery is a small fraction of the retail price. Keshav (2009) concluded that the cost of sending an SMS message was likely to be less than $0.003/SMS, or 2% of the per unit price charged to pay as you go customers (in the USA). Lack of clarity on the true cost of sending an SMS message continues to plague development cost modelling. For example, in a study of an mHealth intervention in Tanzania, Mangone *et al.,* (2016) modelled scenarios using standard SMS charge rates ($0.03/SMS) along with reduced rates ($0.02/SMS and $0.01/SMS), representing negotiated bulk purchasing agreements.

The analysis up to this point has been based on retail prices given in Table 12, which can be interpreted as representing an opportunity cost to the MNO. However, given that MNOs are sending out messages to thousands of users, it could be argued that a bulk SMS price would be more realistic. A price of TZS 25/SMS, for example, would give a contribution margin of 70%. If, on

[14] Halotel was not a partner in 2017 so was not included in the financial report. This is the price given on the website (accessed September 2019).

the other hand, there is no real cost to the MNO of sending SMS messages, then the contribution margin tends towards 100%.

The contribution margin is a measure of how much a product contributes to fixed costs and thereby to profit (once fixed costs are covered). However, it takes no account of any investment made in developing a product, so it cannot be used as a predictive measure of the likely financial viability of developing a product; IRR on investment is widely used for this purpose.

The model indicates that over a four-year period, mobile operators donate texts and USSD sessions totalling a nominal value of £3.9 million. However, over the same time period, the indirect benefit of ARPU sums to a total of £5.1 million, which represents a real financial gain.

The other indirect benefit often credited to VAS is a reduction in churn. However, the quantitative research did not find any evidence that people using the Wazazi Nipendeni service had owned their SIMs for longer. Prepaid markets tend to have high churn rates because consumers are not 'tied in' to contracts. However, in rural or underserved markets where signal coverage is poor consumers often find themselves effectively tied in to a single operator by virtue of being able to access only one network in their geographical location. Therefore, churn tends to be lower in rural areas. The quantitative study was carried out in Iringa, which is a largely rural district, so it is not surprising to find low rates of churn, making it difficult to measure any differences in churn.

7.3 Commercial content provider

The previous section highlighted the financial benefits to MNOs of offering the Wazazi Nipendeni service. This opens up possibilities for the content provider to generate revenue through some kind of revenue share agreement with MNOs, which in turn opens up possibilities for spinning off the content provider as a commercially viable venture. This section considers the financial viability of such a venture as a hypothetical case based on cost data from the mHealth Tanzania-PPP. It is intended to represent the opportunity for replicating an mHealth agency, delivering a service similar to Wazazi Nipendeni, in another country.

The model is based on the following assumptions regarding capital costs:

- Nutrition content developed under the GSMA mNutrition programme is publicly available and open source, so the 'raw' content can be accessed and adapted at no cost.
- There is a real cost associated with localising content. This includes framing the content in messages that are relevant and readily understandable by users, as well as translating content into local languages. In the case of Wazazi Nipendeni, messages are all in written format for disseminating as SMS messages, which avoids the considerable expense of making voice recordings that are needed for outbound dialling and other voice-based systems. Securing government approval of messages is a major cost. Health services tend to be highly regulated, and the experience of Wazazi Nipendeni was that the health ministry was diligent in ensuring that messages were consistent with Government of Tanzania health policies. A good deal of time was spent negotiating the balance between making messages factually accurate yet informal and easy to understand.
- An agency would need to invest in product development, as was the case with Wazazi Nipendeni, which benefited from user experience research and formative evaluation throughout the duration of the mNutrition programme (both of which were funded through the GSMA mNutrition programme).

These costs sum to an investment of £312,500.

The following assumptions have been applied regarding operating costs:

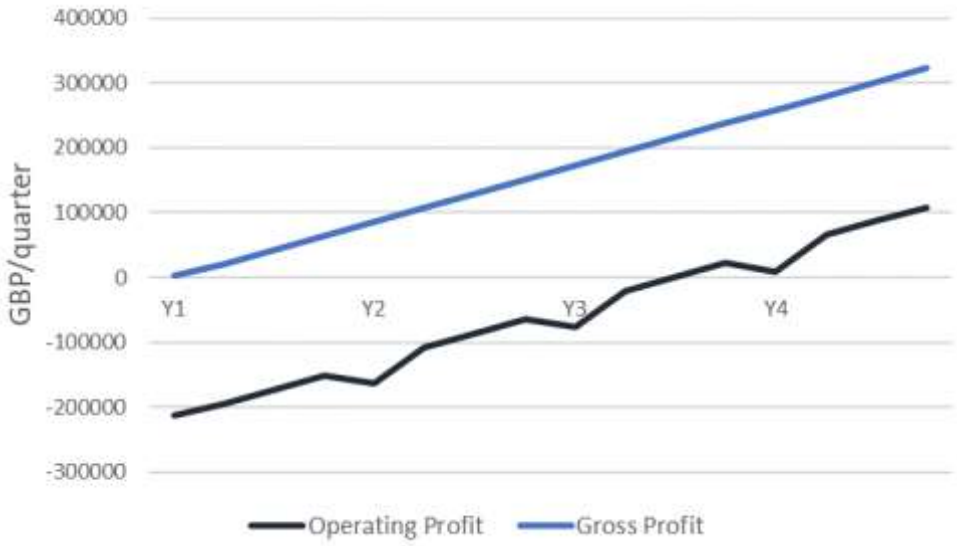
- the fixed operating costs associated with running the mHealth-PPP will not change;
- the content will need to be continually reviewed and revised in order to keep it consistent with developments in government health policy and practice; and
- the agency would need to bulk purchase SMS messages from each client operator. A price of TZS 25/SMS has been assumed.[15].

It is assumed that the hypothetical agency would start from scratch, with no pre-existing users, so the model is based on the profile of active users in Figure 5.

Note that no account has been taken of the in-kind contributions made by government agencies and partner NGOs, both of whom are critical to the success of a service. It can be argued that any government has a vested interest in setting up a service that has a material benefit for health outcomes, and so should forego any potential claims on the cost of the time they invest in mediating on behalf of the service provider. Similarly, it can be argued that the cost of the time invested by partner NGOs, including marketing and registering users, is outweighed by the benefits that the service adds to their field programmes.

At 100% revenue share with MNOs (i.e. all TZS 510/user/month in increased ARPU goes to the content provider), Figure 17 suggests the service would break even at the beginning of Year 3. However, the rate of return on investment over a four-year time horizon would be negative. Given the assumed trajectory of growth in user numbers, much greater revenues are generated if the time scales can be extended. Therefore, at 100% revenue share, the service could generate a positive IRR of 4% over a six-year period. However, rate of return is highly sensitive to revenue share: at 90% revenue share (i.e. the MNO retains only 10% of the increase in ARPU), the IRR turns negative.

Figure 17: Profits – commercial content provider



Source: Authors' own.

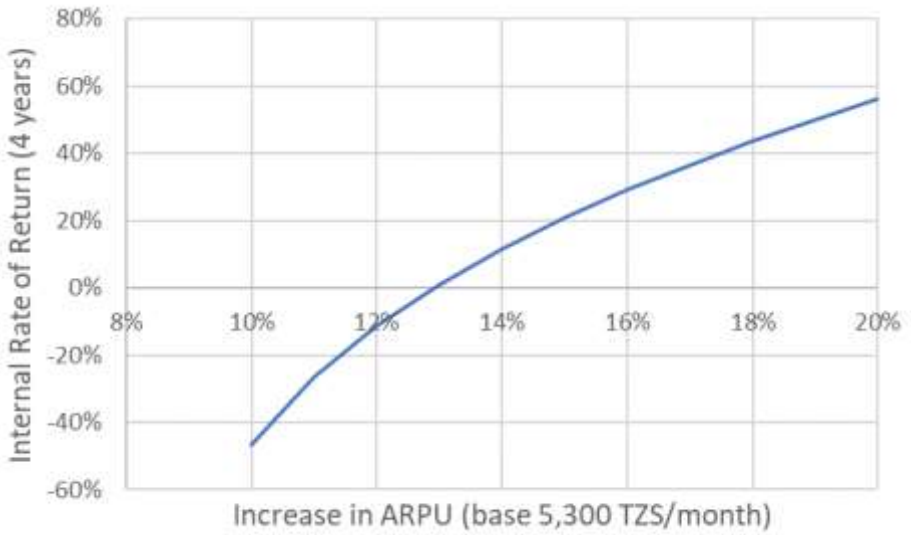
[15] https://sendsms.co.tz/pricing/

Projecting the user growth profile over a further two years would result in a total of 1.3 million active users. The population of Tanzania is 54.2 million (National Bureau of Statistics, 2009) and the average household size is 4.9 (DHS 2014/15), so the number of households across the country is around 11.1 million. 16.2% of the population is under five years of age National Bureau of Statistics, 2013), equating to 8.8 million children under five. If households with children under five had an average of two children both aged under five, then the total number of households with at least one child under five would be 4.4 million (40% of all households). This rough calculation suggests that a user target of 1.2 million Wazazi Nipendeni users may be ambitious but is not impossible to achieve.

As revenue is generated solely from increases in ARPU, the financial viability of such a VAS is highly sensitive to that increase. Up to this point, the analysis has been based on a TZS 510 /month increase in ARPU found by the quantitative study (on a baseline level of ARPU of TZS 5,300 among the control sample). This is based on the effect of the random offer of access to the service identified in the quantitative study (i.e. comparing ARPU between the treatment and control samples). The study found that 66% of the treatment sample and 27% of the control sample reported having received nutrition messages by SMS (in the last two years). The study points out that this method of calculation was likely to underestimate effects, but the authors were unable to calculate more specific effects with any reliability because of inconsistencies in the self-reported access to the Wazazi Nipendeni messages. They concluded that this was because respondents may not have been able to correctly identify the source of nutrition text messages received.

Figure 18 shows how sensitive the financial viability of a service is to this effect size, and that if the effect size was 15% then a service could be financially attractive given a four-year period. Based on the proportions of both samples that reported receiving nutrition message by SMS, the effect size could be double (20%), in which case the return on investment would be highly attractive.

Figure 18: Sensitivity of IRR to increase in ARPU (content provider)

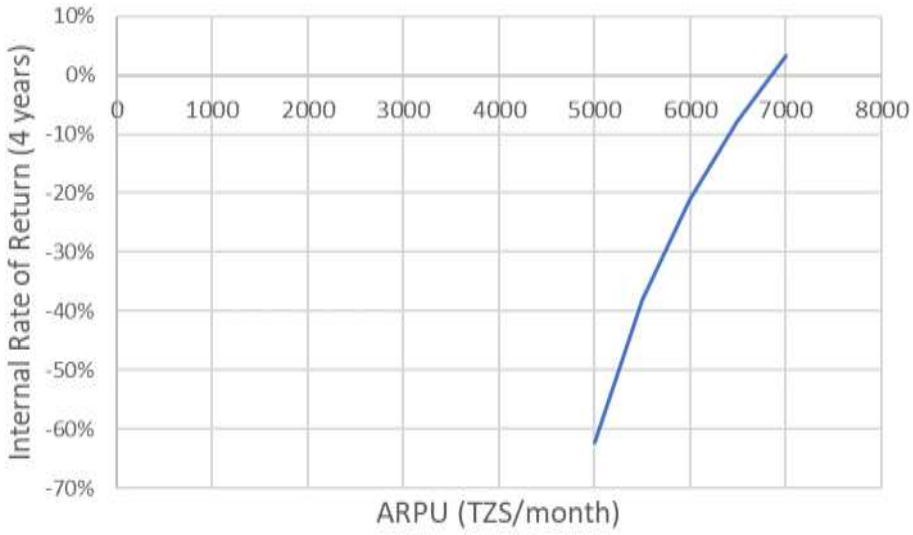


Source: Authors' own.

The large number of active users means that financial indicators are highly sensitive to changes in ARPU. The ARPU found among the control sample was TZS 5,300, which is consistent with published data. Vodacom Tanzania quotes ARPU figures of just over TZS 6,000 (for the year ended March 2019) (Vodacom, 2019), and Airtel has published consolidated data covering five

East African countries[16], quoting an ARPU of $2.4 (December 2018) (Airtel Africa, 2019) (TZS 5,500).[17] The study was conducted in Iringa, a rural district, so measured ARPU levels would be expected to be below that national average. Figure 19 shows that, if all other assumptions are held constant, a service would need to reach a user base with an ARPU over TZS 7,000/month in order to generate a positive IRR over a four-year period. This is unrealistic, given that these reports, as well as industry analysis (BMI Research, 2016), confirm a current trend of declining ARPU. This figure is based on generating revenue from a 10% increase in ARPU.

Figure 19: Sensitivity of IRR to ARPU

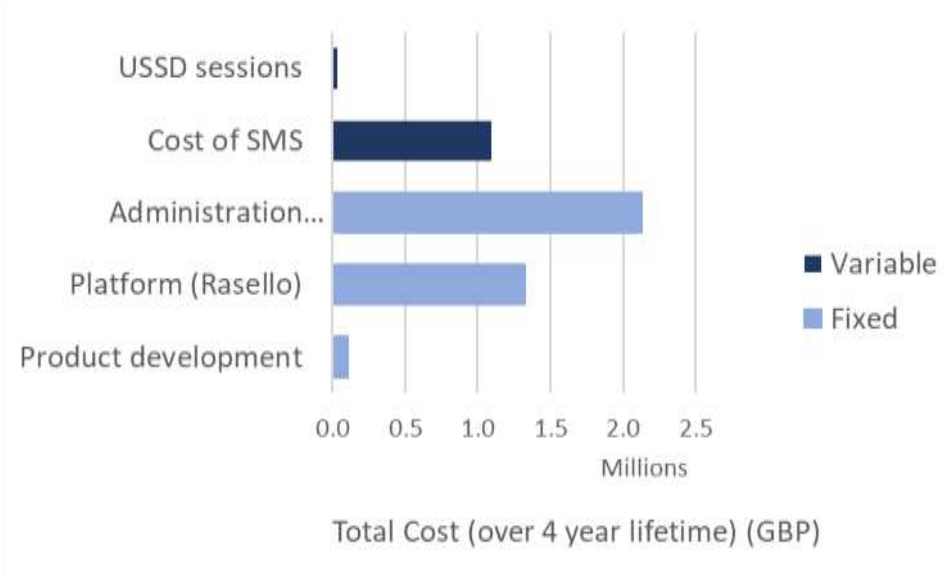


Source: Authors' own.

The breakdown of operating costs over a four-year period shows how the fixed costs associated with running the mHealth Tanzania-PPP, and running the technology platform, are dominant (Figure 20). Analysis shows that central expenses would need to be reduced by 40% in order to produce a positive IRR of 4% over a four-year time period (at 100% revenue share). Again, even at a 90% revenue share the IRR turns negative.

[16] Kenya, Uganda, Rwanda, Tanzania, Malawi, Zambia.
[17] Based on an exchange rate of TZS 2,300/$.

Figure 20: Operating costs (four-year period)



Source: Authors' own.

Consider a scenario in which an independent mHealth VAS provider can develop and deliver an information dissemination service under the following assumptions:

- a 20% reduction in fixed operating costs (staff time and platform costs);
- a nationwide ARPU of TZS 6,000/month;
- a 15% increase in ARPU when consumers use the mHealth VAS;
- a 50% revenue share with MNOs; and
- an SMS price of TZS 25.

These positive assumptions indicate that such as service could generate a positive IRR of 6%, but only over a six-year period.

7.4 In-house MNO service

The quantitative study has shown that consumers using Wazazi Nipendeni spend more money on airtime, estimated to be equivalent to a 10% increase in ARPU. This raises the possibility that an MNO could justify creating an mHealth VAS on the basis of this increased revenue. This section considers a hypothetical business case scenario in which an MNO in another country sets up an mHealth service in-house. The key differences when compared to the commercial content provider scenario considered above is that the MNO would retain all of the additional revenue generated, and there would be no real costs associated with the SMS messages sent.

The model is based on the following assumptions regarding capital costs:

- Nutrition content developed under the GSMA mNutrition programme is publicly available and open source, so the 'raw' content can be accessed and adapted at no cost.
- There is a real cost associated with localising content. This includes framing the content in messages that are relevant and readily understandable by users, as well as translating the content into local languages and securing government approval of messages.

- The MNO would need to invest in product development, as was the case with Wazazi Nipendeni, which benefited from user experience research and formative evaluation throughout the duration of the mNutrition programme.

These costs sum to an investment of £312,500.

The following assumptions have been applied regarding operating costs:

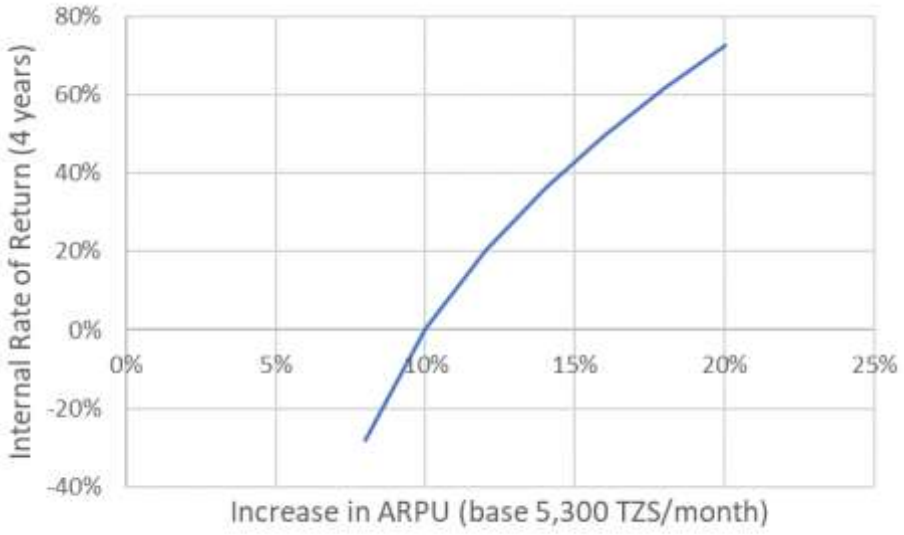
- the fixed operating costs associated with running an mHealth VAS in-house would be the same as for Wazazi Nipendeni;
- the content will need to be continually reviewed and revised in order to keep it consistent with developments in government health policy and practice;
- the MNO would be able to assign the real cost of sending SMS messages and USSD sessions (both have been zero-rated in the analysis); and
- no account has been taken of the in-kind contributions made by government agencies and partner NGOs, both of whom would be critical to the success of such a service.

It is assumed that the hypothetical MNO-based service would start from scratch, with no pre-existing users, so the model is based on the profile of active users presented in Figure 5.

If SMS messages and USSD sessions are zero-rated, then the service would break even in Year 3, although it would not generate enough cash to provide a positive return on investment over a four-year period. Assuming the same trajectory of increasing user numbers, the large number of users (rising to 1.1 million) would generate enough revenue to provide a 19% IRR in a five-year period.

As revenue is generated solely from increases in ARPU, the financial viability of such a VAS is highly sensitive to that increase. Up to this point, the analysis has been based on the TZS 510/month increase in ARPU found by the quantitative study (on a baseline level of ARPU of TZS 5,300 among the control sample), i.e. an effect size of approximately 10%. As described above, this is likely to be an underestimate and the real effect size could be double (a 20% increase in ARPU). Figure 21 shows how sensitive the financial viability of a service is to this effect size, and that even if the effect size was marginally higher than 10% a service could be financially viable in a four-year period. Based on the proportions of both samples that reported receiving nutrition messages by SMS, the effect size could be double (20%), in which case the return on investment would be highly attractive.

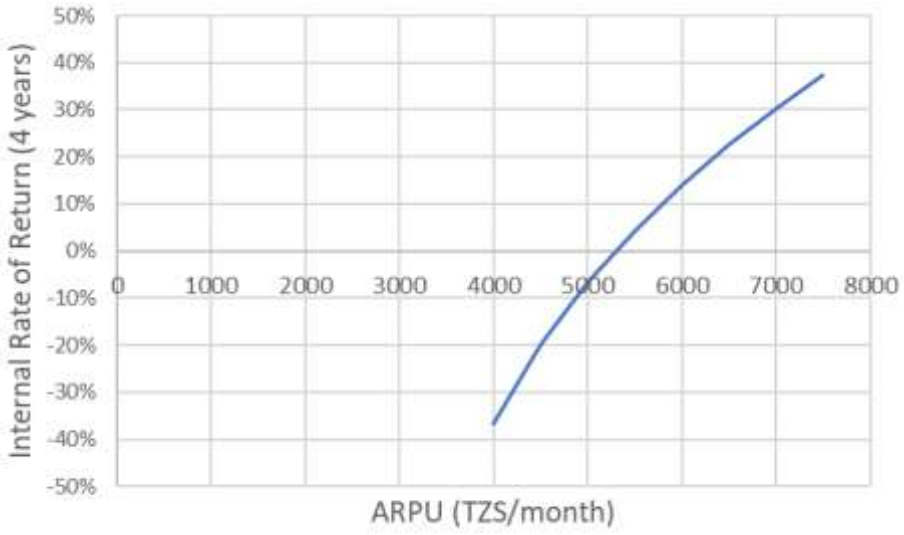
Figure 21: Sensitivity of IRR to increase in ARPU (in-house MNO)



Source: Authors' own.

The large number of active users means that financial indicators are highly sensitive to changes in ARPU. The ARPU found among the control sample was TZS 5,300, which is consistent with published data as described above. Figure 22 shows that a service could generate a positive IRR over a four-year period if offered to a user base with an ARPU of around TZS 6,000, which appears to correspond more or less to the Tanzanian market average. The study was conducted in Iringa, a rural district, so measured ARPU levels would be expected to be below that national average. This figure is based on generating revenue from a 10% increase in ARPU.

Figure 22: Sensitivity of IRR to ARPU (10% effect size)



Source: Authors' own.

Consider a scenario in which an in-house mHealth VAS can be developed and delivered under the following assumptions:

- a 20% reduction in fixed operating costs (staff time and platform costs). An MNO would be quite capable of setting up its own platform, which could reduce real costs still further;

- a nationwide ARPU of TZS 6,000/month; and
- a 15% increase in ARPU when consumers use the mHealth VAS.

These positive assumptions indicate that such a service would be financially attractive, providing an IRR of approximately 70% over a four-year period.

There are non-financial constraints that an MNO would face if implementing an in-house mHealth VAS as described in this section. These relate to relations with government and NGO partners. In the case of Wazazi Nipendeni, TFNC was instrumental in securing approval of the messages. Any government body would need to be seen to be scrupulous in its dealings with a private sector organisation and would be keen to avoid accusations of partisanship. While a government agency might be keen to support an enterprise that is clearly for the public good, it may be more measured in its support for a private sector venture. This depends on the country context; in countries with strong policies on information and communications technology (ICT) and private sector growth, government agencies may be keen to support this kind of venture.

8 Discussion

With no directly measurable effects on the anthropometrics of the children in the study, one could suggest the proposition that DALYs have been averted or saved has limited merit. However, the measurable effects on dietary diversity among children and a discussion of commercial sustainability indicate otherwise.

8.1 Regarding measurable effects

Changes in dietary diversity should result in changes in stunting (Arimond and Ruel, 2004). And with the reduction of stunting, albeit a relatively small amount, there are DALYs averted. Depending on the assumptions applied in the conversion of dietary diversity to DALYs, the study suggests that between 119 (ITT) and 286 (LATE) DALYs have been averted.

Analysis A suggests that the cost of these DALYs **averted is £9,776** (a range of £4,132 to £16,755 (2019 prices)), depending on the assumptions made. Even making the most generous of assumptions, the cost per DALYs averted falls outside the guidelines from the WHO Commission for Macroeconomics and Health (WHO, 2001), which provide the following guideline for thresholds of cost-effectiveness:

> *'An intervention is considered very cost-effective, if the monetary amount spent on the intervention per disability-adjusted life year (DALY) saved is less than the per capita gross domestic product (GDP) for the nation in which the intervention is applied.'*
>
> *'An intervention is considered (moderately) cost-effective, if the monetary amount spent on the intervention per DALY saved is less than three times the per capita GDP.'*

The GDP of Tanzania is £622. This suggests that the intervention is outside the scope of the WHO cost-effectiveness guidelines.

It is also worth noting that Annex C presents an updated literature review, including insights from the Disease Control Priorities report (2018). Using this and a broader consideration of the literature both before and during this study, the cost per DALYs averted cited above is above the upper end of those found, say, in Horton *et al*., (2015) (systematic review, cited in Annex C).

However, having said that, the literature review also found examples of cost-effectiveness where no effect was found, and so one could conclude that the relatively small effect found here is noteworthy. It is also noteworthy that the estimates on change in children's dietary diversity are about half those found in Leroy *et al*., (2017) for a programme that included many other behaviour change communication interventions.

There is no reason to assume the acquisition of new subscribers will not continue. At any given time, there are 5 million children under five whose mothers are a potential beneficiary of the Wazazi Nipendeni service. We have therefore considered a projection for four years and six years at current growth rates resulting in a 1 million and 1.45 million subscriber base, respectively. At this level of subscription, and allowing for increased operational costs, the models actually increase the cost per DALYs averted.

Thus, on the basis of Analysis A, the cost-effectiveness based on DALYs averted does not fit within the guidelines for an acceptable intervention. (We note the many assumptions made;

however, even with the most favourable assumptions, the cost per DALYs averted seems outside the scope of the WHO guidelines.)

Analysis B includes the 'preparation' costs of the GSMA mNutrition programme. It considers the investment made by FCDO to fund the management of the wider mNutrition intervention, and to enhance the Wazazi Nipendeni service with nutritional messaging. Including this investment just pushes the cost per DALYs up and makes the proposition even less attractive when considered through the DALYs lens.

8.2 Commercial sustainability

However, if the increased ARPU is taken into account, then the picture could change dramatically. MNOs are actually benefitting from their contribution to the programme. The contribution margin is well within their expectations of running a profitable business, even if the SMS is costed at the retail rate. (Wazazi Nipendeni reports the MNO contribution to the programme at the SMS retail cost.) In reality, other studies and interviews with MNOs suggest that SMS actually costs the MNO a fraction of the retail price. When this more realistic figure is applied, the contribution margin grows significantly.

In other words, offering SMS for a service like Wazazi Nipendeni is commercially sustainable for all MNOs involved. MNOs are offering zero-rated SMS as part of their corporate social responsibility, and yet they are getting an indirect benefit which gives them revenue which exceeds their 'donation', even when that donation is assigned a retail SMS value.

If the experience of the MNOs is that they get a reasonable contribution margin, and their 'donation' of SMS is more than matched by the increase in ARPU, then from their point of view there is no loss due to an opportunity cost in applying the product. The idea that the product is commercially viable, and could be replicated commercially, suggests that the narrative of the cost-effectiveness centred around DALYs averted could be changed by considering the returns commercially. MNOs could be persuaded to offer the service not based on the number of beneficiaries, but by it being a sound VAS. Whether it then benefits 100 or 100,000 people need not be part of the decision whether to run it or not – it is a viable commercial VAS. The qualitative studies show that the product is appreciated: the perception is that the product or service is valuable. For an MNO it is a viable service to consider offering.

However, the main cost of the programme is shouldered by the PPP arrangement. Staff costs and benefits of the PPP form a considerable part of the ongoing costs. The PPP makes the connections and partnerships through to the Government of Tanzania and other NGOs and coordinates the MNOs. Given that the MNOs are gaining revenue through additional ARPU, Section 7 has shown how such revenue could offset ongoing costs. In a replication of the service in another country, if the MNO led the service, a positive IRR would be relatively easy to achieve. However, that assumes the single MNO can attract a similar number of users. One could argue that the user base for Wazazi Nipendeni is spread over the five participating MNOs, and so any one MNO would get less users. The original vision of GSMA was to identify and support an aggregator which could present a single health and nutritional VAS and sell its services to and coordinate the four MNOs. This is what the PPP does, but it does it as a PPP, not as a private VAS provider. Section 7 shows that if there were revenue share from the MNO of the increased ARPU, then an aggregator could achieve a positive IRR within six years. However, it is unlikely that an aggregator could convince an MNO to share 'increased ARPU' – as an indirect benefit, tracking and allocating the increased ARPU would be a challenge.

Nevertheless, the basic commercial sustainability suggests that the estimates on health returns (DALYs averted) should not only be evaluated against the full cost of the programme. If mechanisms were in place to recover the indirect benefits, it could be possible to envisage a 'costs-effective' service that, while creating minimal behavioural change per se, nevertheless reinforces the general health knowledge of its users, and contributes value alongside other health communications.

8.2.1 Political economy

Although this report states the service could be commercially sustainable if set up by an MNO, or if a positive set of assumptions was applied to an independent content provider, the business modelling component (Scott *et al.*, 2019) documents some key political economy influences on the realities of such a proposition. This perspective is important if the service is to be replicated in another country. However, the landscape of MNOs is rarely a simple one, and the interactions of the private sector and government would come into play in any other country.

For instance, consider the political economy found in Tanzania and documented by Scott *et al*., in the business modelling (Scott *et al.*, 2019).

The telecommunications market in Tanzania has been subject to continued political influence over the duration of the evaluation study. The Electronics and Postal Communications Act (EPOCA) was introduced in 2010 and required telecommunications companies to offer at least 25% of their shares to the public by floating on the Dar es Salam stock exchange, and to do so within three years. The move was resisted by operators and the provision was not enforced at first. However, when a new president was elected (at the end of 2015), the bill was amended in 2016, giving a further three years to make shares available. Vodacom was the first of the operators to comply, listing on the stock exchange in August 2017. At the beginning of 2019, Vodacom was still the only company to have complied.

Tigo in Tanzania is the trading name of MIC Tanzania plc. However, ownership of shares in MIC Tanzania has been in dispute. Two companies owned by a single businessman (Golden Globe International Services Ltd and Quality Group Ltd.) alleged that they bought shares in 2014 entitling them to a 99% stake in MIC Tanzania, although the complex dispute traces its roots back to 2002. Only in July 2018 did the courts rule that MIC Tanzania was the legal owner of Tigo Tanzania. Up until that point, the dispute had meant that Tigo was not able to make provision for any public offering of shares.

Any plans for Airtel to comply with the share offer obligation have also been delayed by a dispute over ownership. Celtel acquired a stake in the Tanzania Telecommunications Company Limited (TTCL, the incumbent operator) in 2001, and the government remained the majority owner. The successful mobile operator, branded as Celtel, was subsequently set up as an independent mobile operator, wholly owned by TTCL. In 2005, the two were legally separated but Celtel retained a 35% stake in TTCL, while TTCL was left with a 40% holding in Celtel. Around the same time (2005), Celtel was purchased by Zain. Zain was then purchased by Bharti Airtel in 2010 and has since been trading as Airtel. In 2017, the government made a claim on the ownership of Airtel, alleging irregularities in the original privatisation process. This argument is not without merit, as one of the problems characterising privatisation processes at this time was a lack of data. The matter has only recently been resolved by Bharti agreeing to make a number of ongoing payments to the government and giving the government further shares to bring its shareholding up from 40% to 49%, leaving Bharti still in control.

Mobile operators have also fallen foul of the judiciary in Tanzania, in line with the president's anti-corruption drive and the targeting of multinational companies in Tanzania's mining and telecoms sectors to address tax evasion. In 2018, the CEOs of Halotel and Zantel were charged with fraud. More recently, in 2019, following the arrest of its CEO, Vodacom Tanzania pleaded guilty to the charge of intending to avoid paying taxes. In April 2019 it made a $2.3 million payment to the government.

Against this backdrop, recasting this 'public good' VAS as a commercially viable proposition would not be straightforward. The political economy would also have to be considered when considering the commercial sustainability of a replicated service.

9 Conclusions and learnings

9.1 The Wazazi Nipendeni Service

The aim of the impact evaluation is to assess the impact, cost-effectiveness, and commercial viability of mNutrition. The evaluation is being conducted by a consortium of researchers from Gamos, IDS, and IFPRI. The team has drawn on a number of methods and interlinked components to gather evidence about the impact of the mNutrition intervention in Ghana, including a qualitative component, a quantitative component, and a business model and cost-effectiveness component.

The design of the cost-effectiveness component relied on there being measurable change in households', women's, or children's dietary diversity. Measures were found of significant measurable change in children's diversity, albeit a relatively weak change. The findings set out above have converted the weak change to DALYs averted, which range from 119 to 286, depending on the assumptions applied.

The analysis of costs has also not been straightforward. The mNutrition service is part of a PPP arrangement within the Tanzanian Government, involving donors, consultants, government staff, and the private sector. While costs for the additional mNutrition content can be identified, the core costs of the PPP are reported as a whole. From interviews, the team have estimated the proportion contribution to staff costs for administration of the service, and the proportion of the Rasello platform. Section 6 shows how the costs used are sensitive to assumptions made regarding proportioning these key costs. Similarly, attributing the costs of sending SMS is not straightforward. Wazazi Nipendeni financial reporting assigns the retail cost of an SMS to in-kind donations made by partner MNOs. This makes sense when reporting to donors, but it is perhaps not the best approach when communicating costs to another party which might be considering replicating the service. Therefore, Analysis A presented the sensitivities of the costs to these key assumptions.

The sensitivity analyses show that the financial model is sensitive to the key cost assumptions made, which are the costs assigned to SMS and the proportion of PPP costs assigned to the service. The quantitative component demonstrated that the service increases household spending on mobile phones, and therefore an increase the ARPU. This is valuable to the MNO. Financial modelling undertaken above concludes that this increased ARPU could possibly be leveraged to produce commercial sustainability. However, such financial viability is highly sensitive to ARPU, and to the effect size on increase in ARPU.

The evaluation as a whole has been able to document and learn from studying the Wazazi Nipendeni VAS. The randomised controlled trial suggests weak behavioural change effects and the cost per DALYs averted are outside the guidelines of the WHO for a cost-effective intervention. However, if commercial sustainability is considered, the possibility of a 'cost-effective' intervention comes to the foreground. The possibility of commercial sustainability suggests that even with marginal beneficial effects, the VAS could be encouraged by policy and financial instruments.

The service was supported by a PPP arrangement with strong involvement of the Government to ensure contextualised quality-of-health information. In theory, such involvement would not be in place with an MNO or private sector aggregator in the lead. The provision of information to the subscriber must be free; however, the indirect benefits suggest that such services can be 'commercially sustainable'. While the original mNutrition programme documents talked about commercial sustainability based on direct revenue generation, the logical framework evolved to talk about imputed benefits to the MNO. The business modelling report discusses how changes in ARPU can benefit the overall MNO bottom line without actually showing as a direct revenue from a

service. An increased ARPU can indeed make the provision of even a free service 'commercially justifiable'.

The remaining challenge is for the revenue, due to indirect benefits, to not only pay for the ongoing costs (by accounting for them in the service budget) but also to pay (at least in part) for the content. Further research is required to identify a mechanism whereby content created using public funds can be paid for through some form of revenue share of the financial benefit yielded by the services that exploit that content.

9.2 Challenges and limitations

The core challenge for the CEA has been the weak measurable impact.

9.2.1 Continuous engagement

Staff turnover. There have been some changes in staff at the partner organisations over the duration of the project, particularly in the MNOs. There are two effects at play. Firstly, there tends to be high turnover rates in the high-tech industry. Personnel who were involved within the MNOs and started engagement with the evaluation team moved on. It was therefore challenging to establish relationships with the MNOs. The second effect is the political economy background, as described in Section 8.2.1. Mergers and restructuring meant that staff were reassigned. In contrast, the mHealth Tanzania-PPP and the Government staff have been remarkably consistent.

Lesson learned: To capture the narrative of an evolving programme such as this, the study team needed to make contact with stakeholders throughout the study – an end-of-programme study alone would not have captured the wider narrative.

9.2.2 Data availability and commercial confidentiality

While the mHealth Tanzania-PPP has been very cooperative in sharing data, the commercial data from the MNOs have not been forthcoming. This is partly due to the sensitivities around commercial data, and partly due to the specific context of Tanzania (and, indeed, partly due to staff turnover). GSMA intended to broker such commercial data but without the leverage of a contractual arrangement it was unable to do so.

Lesson learned: Despite expressions of positive intent to share data, the reality of commercial sensitivities makes it difficult for businesses to share financial data.

9.2.3 Data clarity

The data shared by Rasello on users required considerable cleaning before it could be analysed. There was confusion over the transfer of users from the original system to the Rasello platform, and inconsistencies in opt-out data. A number of assumptions had to be made about the user data in order to create working numbers.

Lesson learned: Commercial collections of data are not by definition accurate. The complexity of sign-in and sign-out processes can create a challenging database.

9.2.4 Cost assumptions

A number of assumptions had to be made regarding the proportion of costs to be assigned to the cost-effectiveness. GSMA did not fund the mHealth Tanzania-PPP directly, and therefore GSMA did not budget for the mHealth Tanzania-PPP and the mHealth Tanzania-PPP did not report to GSMA. The mHealth Tanzania-PPP has cooperated with the study but this has resulted in the need to estimate the proportion of each of the major costs that should be assigned to the VAS.

Lesson learned: Donor reports and financial systems are often not set up to isolate a particular intervention. If an impact study is intended, it would be good to request a breakdown of service costs from the start.

References

Airtel Africa (2019) 'Quarterly report on the results for the fourth quarter and year ended March 31, 2019'.

Aker, J. C., Ghosh, I., and Burrell, J. (2016) 'The promise (and pitfalls) of ICT for agriculture initiatives', in *Agricultural Economics 47*(S1), pp.35-48.

Aker, J.C. and Ksoll, C., 2016. Can mobile phones improve agricultural outcomes? Evidence from a randomized experiment in Niger. Food Policy, 60, pp.44-51.

Alderman, H., Hoddinott, J., and Kinsey, B. (2006) 'Long term consequences of early childhood malnutrition', in *Oxford Economic Papers*, *58*(3), pp.450-474.

Anderson, M. L. (2008) 'Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects', *Journal of the American statistical Association*, 103(484), pp. 1481–1495.

Arimond, M. and Ruel, M.T., (2004) Dietary diversity is associated with child nutritional status: evidence from 11 demographic and health surveys. *The Journal of nutrition*, *134*(10), pp.2579-2585.

Arimond, M., Wiesmann, D., Becquey, E., Carriquiry, A., Daniels, M. C., Deitchler, M., Fanou-Fogny, N., Joseph, M. L., Kennedy, G., and Martin-Prevel, L. E. T. (2010) 'Simple Food Group Diversity Indicators Predict Micronutrient Adequacy of Women's Diets in 5 Diverse, Resource-Poor Settings', *The Journal of nutrition*, *140*(11), pp.2059S-2069S.

Asongu, S. (2015), "The impact of mobile phone penetration on African inequality", International Journal of Social Economics, Vol. 42 No. 8, pp. 706-716.

Barker, D. J. P., Osmond, C., Winter, P. D., Margetts, B., and Simmonds, S. J. (1989) 'Weight in Infancy and Death from Ischaemic Heart Disease', *The Lancet*, *334*(8663), pp.577-580.

Barnett, I., Batchelor, S., Faith, B., Gilligan, D., Gordon, J., Hidrobo, M., Mitchell, B., Palloni, G. and Scott, N. (2020) External Evaluation of Mobile Phone Technology-based Nutrition and Agriculture Advisory Services in Africa: Final Tanzania Mixed Methods Evaluation Report, Brighton: IDS

Barnett, I., Brockerhoff, S., Medardi, D., Gordon, J. and Srivastava, S. (2017) 'External evaluation of mobile phone technology-based nutrition and agriculture advisory services in Africa and South Asia: Mobile phones, nutrition and agriculture in Tanzania: Initial exploratory qualitative study report'. Brighton: IDS

Barnett, I., Faith, B., Mitchell, B., Brockerhoff, S. and Medardi, D. (2019) External evaluation of mobile phone technology-based nutrition and agriculture advisory services in Africa and South Asia Mobile phones, nutrition and health in Tanzania: Qualitative follow-up study report 2019, IDS,: Brighton

Barnett, I., Scott, N., Batchelor, S., and Haddad, L. (2016) 'Dial 'N' for Nutrition? A Landscape Analysis of What We Know About m-Nutrition, m-Agriculture and m-Development', *IDS Working Paper 481*.

Batchelor, S., N. Scott and J. Sharp, (2018) External evaluation of mobile phone technology based nutrition and agriculture advisory services in Africa and South Asia: Mobile phones, nutrition and health in Tanzania Business Model Baseline Report. 2018, IDS, GAMOS: Brighton

Batchelor, S., J. Sharp and N. Scott, (2018) External evaluation of mobile phone technology based nutrition and agriculture advisory services in Africa and South Asia: Mobile phones, nutrition and health in Tanzania: Cost-effectiveness Baseline Report. 2018, GAMOS, IDS: Brighton.

Beegle, K., De Weerdt, J., Friedman, J., and Gibson, J. (2012) 'Methods of household consumption measurement through surveys: Experimental results from Tanzania', *Journal of Development Economics* 98, pp. 3–18.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004) 'How much should we trust differences-in-

differences estimates?', *The Quarterly Journal of Economics,* 119(1), pp. 249–275.

Bhutta, Z. A., Ahmed, T., Black, R. E., Cousens, S., Dewey, K., Giugliani, E., … Shekar, M. (2008) 'What works? Interventions for maternal and child undernutrition and survival', *The Lancet*, *371*(9610), pp.417-440.

Black, R. E., Victora, C. G., Walker, S. P., Bhutta, Z. A., Christian, P., De Onis, M., … Uauy, R. (2013) 'Maternal and child undernutrition and overweight in low-income and middle-income countries', *The Lancet 382*(9890), pp.427-451.

Blackburn, J., Stanojevic, R., Erramilli, V., Iamnitchi, A. and Papagiannaki, K., (2013) September. *Last call for the buffet: economics of cellular networks.* In Proceedings of the 19th annual international conference on Mobile computing & networking (pp. 111-122). ACM.

BMI Research (2016) 'Tanzania Telecommunications Report Q3 2016'.

Christian, P., Lee, S. E., Angel, M. D., Adair, L. S., Arifeen, S. E., Ashorn, P., … Black, R. E. (2013) 'Risk of childhood undernutrition related to small-for-gestational age and preterm birth in low- and middle-income countries', *International Journal of Epidemiology*.

Cole, S. A. and Fernando, A. (2012) 'Mobile'izing Agricultural Advice: Technology Adoption, Diffusion and Sustainability', *SSRN*.

Courtois, P. and Subervie, J. (2014) 'Farmer bargaining power and market information services', *American Journal of Agricultural Economics*, *97*(3), pp.953-977.

Danquah, M. and Iddrisu, A.M., 2018. Access to mobile phones and the wellbeing of non-farm enterprise households: Evidence from Ghana. Technology in Society, 54, pp.1-9.

De Onis, M., Dewey, K.G., Borghi, E., Onyango, A.W., Blössner, M., Daelmans, B., … Branca, F. (2013) 'The World Health Organization's global target for reducing childhood stunting by 2025: Rationale and proposed actions', *Maternal and Child Nutrition*, *9*, pp.6-26.

Demographic and Health Surveys (2016) *[Datasets]*, ICF International, Calverton, Maryland.

Dewey, K. G., Cohen, R. J., Arimond, M., and Ruel, M. T. (2006) 'Developing and Validating Simple Indicators of Complementary Food Intake and Nutrient Density for Breastfed Children in Developing Countries. The food and Nutrition Technical Assistance (FANTA) Project', *Academy for Educational Development*.

Dunet, D. (2012) CDC coffee break: introduction to economic evaluation. Atlanta. Centers for Disease Control.

Fafchamps, M., and Minten, B. (2012) 'Impact of SMS-based agricultural information on Indian farmers', *World Bank Economic Review*, *26*(3), pp.383-414.

Fisher, R. A. (1935) 'The logic of inductive inference', *Journal of the Royal Statistical Society*, 98(1), pp. 39–82.

Flax, V. L., Negerie, M., Ibrahim, A. U., Leatherman, S., Daza, E. J., and Bentley, M. E. (2014) 'Integrating Group Counseling, Cell Phone Messaging, and Participant-Generated Songs and Dramas into a Microcredit Program Increases Nigerian Women's Adherence to International Breastfeeding Recommendations', *The Journal of Nutrition*, *144*(7), pp.1120-1124.

Food and Agriculture Organization of the United Nations (2018) *FAOSTAT Statistics Database*, Food and Agriculture Organization of the United Nations, Rome.

Gelli, A., Margolies, A., Santacroce, M., Sproule, K., Theis, S., Roschnik, N., Twalibu, A., Chidalengwa, G., Cooper, A., Moorhead, T., Gladstone, M., Kariger, P., and Kutundu, M. (2017) 'Improving child nutrition and development through community-based childcare centres in Malawi - the NEEP-IE study. Study protocol for a randomized controlled trial' in *Trials*, 18(1), p. 284.

Gilligan, D., Hidrobo, M., Ledlie, N., and Palloni, G. (2018) 'Mobile Phones, Nutrition, and Health in Tanzania: Quantitative Baseline Report'. External evaluation of mobile phone technology-based nutrition and agriculture advisory services in Africa and South Asia, IFPRI/IDS/OPML

Gilligan, D, Hidrobo, M, Palloni, G and Tambet, H, (forthcoming, 2020) 'Mobile phones, nutrition, and health in Tanzania: Quantitative Endline Report'. External evaluation of mobile phone technology-based nutrition and agriculture advisory services in Africa and South Asia, IFPRI/IDS/OPML

Gluckman, P. D. and Hanson, M. A. (2004) 'Living with the past: Evolution, development, and patterns of disease', *Science*, *305*(5691), pp.1733-1736.

Greevy, R., Silber J. H., Cnaan, A., and Rosenbaum, P. R. (2004) 'Randomization inference with imperfect compliance in the ACE-inhibitor after Anthracycline randomized trial', *Journal of the American Statistical Association* 99(465), pp. 7–15.

Hildebrandt, N., Nyarko, Y., Romagnoli, G., and Soldani, E. (2015) 'Price information, inter-village networks, and "bargaining spillovers": Experimental evidence from Ghana', *Working Paper*.

Hoddinott, J., Behrman, J. R., Maluccio, J. A., Melgar, P., Quisumbing, A. R., Ramirez-Zea, M., … Martorell, R. (2013) 'Adult consequences of growth failure in early childhood', *American Journal of Clinical Nutrition*, *98*(5), pp.1170-1178.

Huybregts, L., Becquey, E., Zongrone, A., Le Port, A., Khassanova, R., Coulibaly, L., Leroy, J. L., Rawat, R., and Ruel, M. T. (2017) 'The impact of integrated prevention and treatment on child malnutrition and health: the PROMIS project, a randomized control trial in Burkina Faso and Mali', *BMC Public Health* 17(237).

Imbens, G. W. (2015) 'Matching Methods in Practice: Three Examples', *Journal of Human Resources*, *50*(2), pp.373-419.

Imbens, G. W., and Rosenbaum, P. R. (2005) 'Robust, accurate confidence intervals with a weak instrument: quarter of birth and education', *Journal of the Royal Statistical Society*, Series A 168(Part 1), pp. 109–126.

Imbens, G. W., and Rubin, D. B. (2015) 'Causal inference: For statistics, social, and biomedical sciences an introduction', in *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press.

Institute for Health Metrics and Evaluation (2018a) 'Global Burden of Disease Study 2017. Results'. Seattle, United States.

Institute for Health Metrics and Evaluation (2018b) 'GBD 2017 Online Tools Overview'. Seattle, United States.

International Labour Organization (2018) *ILOSTAT database*. [Data retrieved 18 April 2019].

Jiang, H., Li, M., Wen, L. M., Hu, Q., Yang, D., He, G., Baur, L.A. and Qian, X. (2013) 'Effect of a short message service intervention on infant feeding practices in Shanghai, China: a prospective, community-based, controlled study', *The Lancet Global Health*, *382*, p.4.

Keshav S. (2009) *Cell Phone Text Messaging Rate Increases and the State of Competition in the Wireless Market*. Senate Subcommittee on Antitrust, Competition Policy and Consumer Rights.

Kim, S.S., Nguyen, P.H., Yohannes, Y., Abebe, Y., Tharaney, M., Drummond, E., Frongillo, E.A., Ruel, M.T., and Menon, P. (2019) 'Behavior Change Interventions Delivered through Interpersonal Communication, Agricultural Activities, Community Mobilization, and Mass Media Increase Complementary Feeding Practices and Reduce Child Stunting in Ethiopia', *The Journal of Nutrition*.

Kleinbergen, F. and Paap, R. (2006) 'Generalized Reduced Rank Tests Using the Singular Value Decomposition', *The Journal of Econometrics.*

Labrique, A., Vasudevan, L., Kochi, E., Fabricant, R., and Mehl, G. (2013) ' mHealth innovations as health system strengthening tools: 12 Common Applications and a Visual Framework', *Global Health: Science and Practice*, *1*(2), pp.160-171.

Leroy, J., Olney, D., Bliznashka, L., and Ruel, M. (2015) 'The impact of Tubaramure, a food-assisted integrated health and nutrition program in Burundi, on household food consumption and maternal and child dietary diversity', *The FASEB Journal*, *29*(1_supplement), pp.39-7.

Mangone, E.R., Agarwal, S., L'Engle, K., Lasway, C., Zan, T., Van Beijma, H., Orkis, J. and Karam, R., (2016) Sustainable cost models for mHealth at scale: modeling program data from m4RH Tanzania*. PloS one*, *11*(1), p.e0148011.

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

MoHCDGEC [Tanzania Mainland], Ministry of Health (MoH) [Zanzibar], National Bureau of Statistics, Office of the Chief Government Statistician, and ICF (2016) 'Tanzania Demographic and Health Survey and Malaria Indicator Survey (TDHS-MIS) 2015-16', MoHCDGEC, MoH, NBS, OCGS, and ICF, Dar es Salaam, Tanzania, and Rockville, Maryland, USA.

Moursi, M. M., Arimond, M., Dewey, K. G., Treche, S., Ruel, M. T., and Delpeuch, F. (2008) 'Dietary Diversity is a Good Predictor of the Micronutrient Density of the Diet of 6- to 23-month-old Children in Madagascar', *Journal of Nutrition*, *138*(12), pp.2448-2453*.*

National Bureau of Statistics (2019) 'Tanzania in Figures 2018', The United Republic of Tanzania.

National Bureau of Statistics (2013) 'Population Distribution by Age and Sex', National Bureau of Statistics, Ministry of Finance, Dar es Salaam, and Office of Chief Government Statistician, President's Office, Finance, Economy and Development Planning Zanzibar.

Neumann, P.J., (2009) Costing and perspective in published cost-effectiveness analysis. Medical care, pp.S28-S32.

Nglazi, M. D., Bekker, L. G., Wood, R., Hussey, G. D., and Wiysonge, C. S. (2013) 'Mobile phone text messaging for promoting adherence to anti-tuberculosis treatment: a systematic review protocol', in *Systematic Reviews*, *2*(1), p.6.

Pelletier, D. L., Frongillo, E. A., Schroeder, D. G., and Habicht, J. P. (1995) 'The effects of malnutrition on child mortality in developing countries', *Bulletin of the World Health Organization*, *73*(4), p.443.

Population and Housing Census (2012) 'Iringa Region: Basic Demographic and Socio-Economic Profile', technical report, National Bureau of Statistics of the Government of Tanzania.

Poushter, J., Bell, J., Carle, J., Cuddington, D., Deane, C., Devlin, K., … Zainbulbhai, H. (2015) 'Cell Phones in Africa : Communication Lifeline', *Pew Research Centre*.

Rokicki, S., Cohen, J., Salomon, J. A., and Fink, G. (2017) 'Impact of a text-messaging program on adolescent reproductive health: A cluster-randomized trial in Ghana', *American Journal of Public Health*, *107*(2), pp.298-305.

Rosenbaum, P. (2002) 'Covariance adjustment in randomized experiments and observational studies', *Statistical Science* 17(3), pp. 286–327.

Saaka M, Oladele J, Larbi A, Hoeschle-Zeledon I. (2017) 'Dietary Diversity Is Not Associated with Haematological Status of Pregnant Women Resident in Rural Areas of Northern Ghana', *J Nutr Metab*, 8497892. doi:10.1155/2017/8497892.

Scott, N., Batchelor, S., Jones, T. (2020) External evaluation of mobile phone technology-based nutrition and agriculture advisory services in Africa and South Asia Mobile phones, nutrition, and agriculture in Tanzania: Business Modelling Endline Report.

Small, D. S., Ten Have T. R., and Rosenbaum, P. R. (2008) 'Randomization inference in a group-randomized trial of treatments for depression: Covariate adjustment, noncompliance, and quantile effects', *Journal of the American Statistical Association 103*(481), pp. 271–279.

Staiger, D., and Stock, J. H. (1997) 'Instrumental variables regression with weak instruments', *Econometrica* 65, pp. 557–586.

Stock, J. H. and Yogo, M. (2005) 'Testing for Weak Instruments in Linear IV Regression', in *Identification and Inference for Econometric Models,* Cambridge University Press, New York.

Svensson, J. and Yanagizawa, D. (2009) 'Getting prices right: The impact of the market information service in Uganda', *Journal of the European Economic Association*, *7*(2-3),

pp.435-445.

Tanzania National Bureau of Statistics (2012) *Tanzania Household Budget Survey (HBS) 2011–2012.*

TFNC (2014) ‘Tanzania National Nutrition Survey 2014’.

Victora, C. G., Adair, L., Fall, C., Hallal, P. C., Martorell, R., Richter, L., and Sachdev, H. S. (2008) 'Maternal and child undernutrition: consequences for adult health and human capital', *The Lancet*, *371*(9609), pp.340-357.

Vodacom (2019) Preliminary consolidated financial results for the year ended 31 March 2019.

WHO (2008) 'Indicators for Assessing Infant and Young Child Feeding Practices: Part 1 Definitions', technical report, WHO.

WHO (2001) Commission on Macroeconomics and Health. Macroeconomics and health: investing in health for economic development. Geneva: WHO, 2001.

World Bank (2018) *World Development Indicators Online*, The World Bank, Washington, DC.

Annex A Terms of reference

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 10 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 10 countries in Africa (Cote d'Ivoire, Ghana, Kenya, Malawi, Mozambique, Nigeria, Rwanda, Tanzania, Uganda and Zambia) and 4 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 3 million households in 10 countries in Africa and 4 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 3 million poor people and community health workers across 10 SSA and 4 Asian countries;

- Launch and scaling of mobile phone based health, nutrition and agricultural advisory services targeted to 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 SSA; 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 pregnancies; infant and young child feeding practice; and micro-nutrient supplementation to children at risk (i.e. Vitamin A, Zinc and Oral Rehydration Solution (ORS)).

mNutrition has started implementation from September 2013. For the 2 countries selected for the impact evaluation (Tanzania and Ghana), mobile network operators and content providers have been identified through a competitive process 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 till 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% 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 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?

Further evaluation questions related to other aims of mNutrition will be addressed in at least 1 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 Organisation) 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 specific target audience in the 14 countries. The adapted content and nature of messages is expected to vary across specific 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 specific 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 service 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 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 3 million people will be reached through mNutrition; including 2 million for nutrition sensitive agriculture advisory messages in 4 Asian and at least 2 African countries and about 1 million beneficiaries for mobile phone based nutrition services in 10 countries in SSA.

The evaluation contract period will be September 2014 to 31st 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 generalization 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 6 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 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 mNutrition service 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 end-lines, 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 subcontract 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 2 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 learnt about mobile phone based extension approaches in nutrition and agriculture.

Deliverables

The Evaluator will deliver the following outputs[18]:

During the design and study inception phase of maximum 4 months:

- A publishable landscape analysis report highlighting lessons learnt from existing initiatives on mobile phone based advisory services related to nutrition and agriculture by month 4;
- An updated work plan 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:

- 4 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

[18] Exact timeframe of deliverables will be agreed on during the design phase as appropriate.

  - Two Baseline quantitative reports submitted by April 2017
  - Two Baseline qualitative reports submitted by February 2017
  - Two Cost-effectiveness reports 1 submitted by March 2017
  - Two Business Model reports 1 submitted by March 2017
  - Two Mixed Methods Baseline reports completed by September 2017
  - Two Midline qualitative reports submitted by March 2018
- 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 report 2 submitted by July 2019
- Two Business Model report 2 submitted by July 2019
- Two Evaluation reports submitted by October 2019
- At least 1 article, based on the findings from the country evaluation reports, published in a research journal;
- A shared lesson learnt 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 sufficiently 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 Advisory Group will meet on a bi-annual basis and comprises of representatives of DFID, NORAD and GSMA representatives and independent technical experts. The Evaluator will be managed by DFID on behalf of the mNutrition Advisory Group. The Evaluator will work closely with the mNutrition service 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 data sets.

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 the Advisory Group. The Evaluator will report directly to DFID who will manage the evaluation on behalf of the mNutrition Advisory Group. The main point of contact for technical matters is Louise Horner, Livelihoods Adviser and Hugh McGhie, Deputy Programme Manager for all other project related issues. The mNutrition Advisory

Group will be the arbiter of any disputes between the evaluation function and the overall programme implementation.

At the end of each 6 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 6 monthly reports outlined above, the Evaluator will provide information to feed into the DFID Annual Review of mNutrition. The 6 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 HMG 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 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 out 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, 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 briefing as outlined above. Travel advice is also available on the FCO 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 have 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.

Annex B Stakeholder interview list

Key stakeholders	
GSMA	Natalia Pshenichnaya (head of mNutrition)
GSMA	Alexander Roche (business intelligence mHealth)
GSMA	Diana Sang (representative for Tanzania)
GSMA	Willie Ngumi
GSMA	Mojca Cargo
GSMA	Kate Zechner
GSMA	Tobias Wacker (user research)
GAIN	Kyla Stockdale (Senior Programme Manager)
mHealth Tanzania-PPP	Muttah Saulo (monitoring and evaluation)
mHealth Tanzania-PPP	Ms Janita Ferentinos
mHealth Tanzania-PPP	Peter Maro (team lead)
mHealth Tanzania-PPP	Said Ali Karume (IT adviser)
mHealth Tanzania-PPP	Mr Francis Chiduo
Cardno	Violet Ketani (programme manager, P4 Project)
MoHCDGEC	Walter Ndesanjo (ICT officer, ICT Unit)
MoHCDGEC	Leyla Bungire (program officer – nutrition)
MoHCDGEC	Hermes Sotter (Ministry of Health eHealth)
MoHCDGEC	Dr Azma Simba (Ministry of Health)
MoHCDGEC	Dr Joyceline Kaganda
MoHCDGEC	Ms Valeria Milinga
MoHCDGEC	Mr Peter Kaswahili
MoHCDGEC	Deus Gwanchele
TFNC	Ms Victoria Kariathi
TFNC	Elizabeth Lyimo (Maternal, infant, and young child nutrition team)
TFNC	Mrs Maria Msangi
Nutrition International	Mr Daniel Nyagawa
Viamo	Hannah Metcalfe
FHI 360 – US Agency for International Development Tulonge Afya	Mr Marcos Mzeru
Totohealth	Felix Kimaru (CEO)
IMA World Health	Mr Joseph Mugyabuso
COUNSENUTH	Ms Belinda Liana
Tigo	Ms Halima Okash
United Nations Children's Fund	Ms Tuzie Ndekie

Annex C mHealth comparable cost-effectiveness

The baseline cost-effectiveness report included a literature review, to identify recent learning on the cost-effectiveness of mHealth projects. In anticipation of some effectiveness in relation to the primary outcomes, the team updated this literature review to ensure the final report would be based on the latest learnings on cost-effectiveness. It is worth noting that from 2016 to the current date there have been a number of significant studies on cost-effectiveness, and there has also been a community-wide commentary on, and rethink of, CEA.

C.1 What are the most significant studies for mNutrition?

LeFevre *et al*., (2018) examine how mobile SMS maternal health information messaging affects utilisation rates for maternal and child health services in South Africa. This is a similar intervention to mNutrition mHealth in terms of the delivery of health advice to pregnant women and new mothers. However, the outcomes tracked are not direct health outcomes: instead, the authors measure the number of antenatal care visits and childhood immunisations. Over two years, they track changes in these utilisation rates compared to a control arm where no SMS messages were sent.

Their CEA is interesting for two reasons. Firstly, there is a lack of significance of their measured outcomes. Due to a small number of participants, incomplete data, and a high baseline utilisation of the services, the study was not powered to indicate a significance to the trend in the increase of utilisation they observed. Nonetheless, the authors use these utilisation trends in the LiST tool to calculate DALYs averted as part of their economic analysis.

Secondly, rather than analysing the intervention for the two years it was running, the authors shifted to constructing a model to calculate outcomes and costs if the intervention was scaled up over five years, to reach 60% of women in Gauteng province in South Africa. The model uses the changes in utilisation rates found in the study, the monthly uptake trends observed to model the roll-out, and cost data from the actual intervention. When discussing the limitation to their analysis, the authors acknowledge that, initially, they wanted to base their analysis on the primary data with a two-year analytic time horizon. They state that data collection was challenging, leaving their data small and incomplete. Instead, therefore, they shifted to a model-based analysis to more rigorously capture the uncertainty in their data, and to investigate scale. It is interesting that results that could not be determined to be significant were used as part of the model, but there is a sensitivity analysis which investigated how the outcomes might vary with changes in the uncertain parameters.

As would be expected intuitively, the cost-effectiveness of the intervention improves over the five-year period, with more people reached and more health benefits assumed to have occurred (through the LiST tool).

Willcox *et al.,* (2019) put forward an *ex-ante* scale-up analysis of an mHealth intervention including maternal messaging, based on data from an intervention in a district in Ghana. There are two parts to the intervention: mobile midwives send voice messages to pregnant and new mothers, and a digital platform for health workers enables them to track and record women's health service utilisation. The mobile midwife aspect of this intervention in comparable to mNutrition mHealth's programme.

The scale-up analysis is based on data from an independent evaluation of the mobile technology for community health (MOTECH) programme, which encompasses the two aspects as described above, in one district in Ghana. Following up with previous publications, LeFevre *et al*., (2017) give

insights into the challenges the mobile midwife programme faced (although this may not be the same district referred to in Willcox *et al*., (2019). Challenges with the technical functionality greatly affected the intervention, as 25% or less of the messages were received by pregnant women. LeFevre *et al*., (2017) think that delivery was hampered by poor network, phones being off, call congestion, and issues with the MOTECH technical system. Once messages had been delivered, 80% were listened to by pregnant women, but this listening rate decreased postpartum – at 6–12-month postpartum, less than 6% of mothers enrolled in the trial listened to at least one message. This detail highlights that technical issues in message delivery, and then whether the women actually listened to the messages, were key issues in how the mobile midwife programme worked. They also state that they initially sought to monitor care-seeking practices of the women in response to the messages, but issues with data collection meant this was not possible.

Referring back to Willcox *et al.,* (2019), data from one district are used to forecast the cost-effectiveness of scaling up MOTECH to 170 districts in Ghana. The data in question are stated to have come from an independent evaluation which could not be found, and so it is not entirely clear how they relate to the findings of LeFevre *et al*., (2017) and the issues found with the mobile midwife programme. Nonetheless, a 10-year analytic time horizon is used, with cost data and outcome data based on the evaluation of the single district (in comparison to a control district). The outcomes that are found to be significant are increases in skilled birth attendants, facility delivery, and rates of measles vaccination, and so these are used in the LiST tool to project control and implementation scenarios forward. Willcox *et al.,* carry out both a deterministic and probabilistic calculation for the cost-effectiveness analyses and find the results to be within the same order of magnitude.

Patel *et al*., (2018) have authored a study of another mobile phone-based health intervention, where women received a weekly phone call and daily text messages, and a counsellor was on hand to speak to at any time. The advice was on antenatal care, iron folic supplements, maternal nutrition, IYCF practices, immunisations, and breastfeeding problems, and the aim was to encourage better breastfeeding practices. The study occurred in India. The CEA focused on whether the scheme improved exclusive breastfeeding rates at six months of age. Rather than analysing the intervention, a different approach to assess the cost-effectiveness was used: a 'boot-strap' method involved resampling to 100,000 observations. It is not clear why this was done, as opposed to just assessing the intervention, as there were just over 500 women in the control and treatment arms. The trial showed positive results in terms of breastfeeding practices, but the cost-effectiveness is not reported succinctly, which makes comparison with other inventions difficult: 'For 5603 rupees/127 USD, a 50% improvement in exclusive breastfeeding at 6 months can be achieved.'

The fourth study which is highlighted as relevant to mNutrition is Nisbett *et al.,* (2016), which is an impact evaluation of a FCDO programme to improve nutrition in Bangladesh. This is of relevance to mNutrition because it is an evaluation of a FCDO project that has some similarities to mNutrition, with the evaluation commissioned by and relevant for FCDO, although there are differences in setting and intervention types. It is also of interest because of the lack of significant outcomes that were measured, and what direction the evaluation took due to this.

The project aimed to improve nutritional outcomes for children, mothers, and adolescent girls, by integrating nutrition-specific interventions into three existing programmes that provide livelihood support. The aspect of integrating nutrition activities into existing programmes is similar to the way nutrition messages are integrated into existing mAgri and mHealth initiatives in mNutrition. The differences are that there is no element of mobile phone advice. Instead, the three existing livelihoods programmes cover grants, infrastructure, and education support on a range of topics

and areas: accessing markets, banking, keeping children in school, capacity building in business, technology transfer and cash transfer, and improving infrastructure such as sanitation.

The nutritional aspect that was integrated into these three programmes targeted lactating mothers, children under five years, and adolescent girls. The invention was behaviour change communications delivered by community nutrition workers, on topics such as breastfeeding, complementary feeding, and hygiene. In addition, micronutrients were provided, deworming was encouraged, and community discussions were facilitated on the needs of adolescent girls.

One aim of the evaluation was to assess quantitatively the impact of the combined livelihood and nutrition programme on the nutritional status of children under two years old. This would be compared with the livelihood programme only. Other outcomes were also of interest: mothers' knowledge and attitudes, IYCF practices, supplement taking, dietary diversity of mothers and children, BMI of mothers, water, sanitation and hygiene (WASH) practices, and women's empowerment.

The CEA plan shows that the authors planned to assess stunting as the primary outcome, and to convert this outcome to DALYs averted. They also state that they followed the FCDO value for money approach, exploring economy, efficiency, effectiveness, and equity (see value for money report). The aim was to back up the quantitative data with understanding about outcomes and the programme effectiveness obtained through qualitative data-gathering processes.

Unfortunately, the authors found that few outcomes experienced a significant change. There was no significant change in: complementary feeding practices, dietary diversity of children, intake of food from animal sources, WASH practices, latrine access, child illness, women's empowerment, child nutritional indicators (stunting, wasting, height-for-age, weight-for-height), household dietary diversity, or women's BMI. Positive changes were found in relation to: antenatal care attendance, participation in feeding programmes, and the fact that women had more voice in decisions regarding where they could go alone.

The reasons for the lack of impact are discussed by the authors, and it is acknowledged that the implementation of home counselling by the community nutrition workers was particularly challenging. There were very few actual visits, which meant there was limited time with mothers. This was caused by high caseloads and long travel times. The authors also discuss the insufficient integration between the management structures for the livelihoods and nutrition programmes.

Without a measurable health impact, assessing cost-effectiveness as planned through the quantification of DALYs averted was not possible. Therefore, the approach was adapted: 'the central research questions could still be answered usefully via an adapted approach' (Nisbett *et al.*, 2016). This pivot was agreed with FCDO. The assessment focused more on economy, efficiency, effectiveness, and equity, and on the learning and policy lessons from the findings. The evaluation conclusion was that the project was not performing well: although on the plus side it was low-cost, it was not delivering real change.

The final study of interest is a study of the Working to Improve Nutrition in Northern Nigeria (WINNN) project, evaluated by OPM (Shehzad *et al*., (2017)), main and summary cost-effectiveness report. CEA is used for two aspects of the project – the IYCF counselling component, and the community-based management of malnutrition component, which focuses on severe acute malnutrition. The IYCF component is potentially comparable to the IYCF component in mHealth. WINNN was a six-year FCDO-funded project, and the evaluation appears thorough, carefully laid out, and contains the details of all the calculations performed. The interventions occurred on the scale of local government areas, and the IYCF activities were conducted through the community or health facilities.

The IYCF CEA used a 'difference-in-difference' methodology. This means that the costs and outcomes are assessed at baseline and endline, for both treatment (WINNN programme present) and control (WINNN programme not present) scenarios. For each scenario, the difference in outcomes and costs between baseline and endline are calculated. Then the scenarios are compared, and the differences found between the treatment and control outcome and cost metrics. Finally, the ratio is taken of these differences-in-difference measures (outcomes/costs) to obtain the incremental cost-effectiveness ratio. The authors took this approach because IYCF interventions were also available in other areas where WINNN was not operating, so it was necessary to take account of the changes that might be occurring in the control group.

A decision tree model was used to map out the potential pathways of exposure to IYCF interventions and the outcomes. The number of children on each tree branch/pathway was based on data from the authors' endline survey. The measured outcome was breastfeeding rates, and these were used in the LiST tool to project mortality, which was in turn converted to DALYs averted through an assumption about the number of DALYs lost due to premature death. The decision tree model compares two scenarios – WINNN in action in the local area, or WINNN not – and the total costs and total mortality outcome are evaluated for each using the decision tree. This is done for baseline and endline and the difference-in-differences method was applied as discussed in the above paragraph.

The WINNN evaluation took two perspectives for the CEA: health service provider (which included government costs, WINNN programme costs, and health facility costs), and societal (which added on caregiver costs and community volunteer costs). There is some discussion about cost-effectiveness thresholds, and the authors check their cost-effectiveness estimates against both the WHO guidelines and a threshold based on opportunity cost proposed by Woods *et al*., (2016).

The main report makes reference to a forthcoming/separate value for money report, which the cost-effectiveness reports will feed into – however, this could not be found.

C.2 The Disease Control Priorities

The *Disease Control Priorities* third edition (DCP3) was written by the Disease Control Priorities Network through a nine-year project starting in 2010, funded by the Bill & Melinda Gates Foundation. It is made up of nine volumes and was released over three years, from 2015 to 2018. Before that, the Disease Control Priorities Project released DCP1 in 1993 and DCP2 in 2006. The reports provide a review of the evidence of cost-effective interventions, as well as reviewing the burden of disease in low- and middle-income countries. The aim of the work is to 'promote and support the use of economic evaluation for priority setting at both global and national levels'.

DCP3 seems an important document to be familiar with because many cost-effectiveness metrics in common use are collated in it and sourced from it. However, it should be noted these are presented as general metrics for interventions meant in a broad sense, and that the calculations and estimates are not necessarily transparent (discussed below).

DCP3 provides a broad but thorough overview of relevant health issues and interventions that have relevance to mNutrition. There is no mention of mobile phone interventions, apart from a brief discussion about how these can support community health workers.

Chapter 17 in DCP3 is dedicated to the cost-effectiveness of interventions for RMNCH. The cost-effectiveness metrics are given as ranges and have been compiled from other studies. All metrics are incremental cost-effectiveness ratios (they are calculated relative to a status quo, control arm, or similar). Four interventions are related to nutrition. Three are extracted here. The fourth is a 'Comprehensive nutrition package' referenced to all interventions is a study called Lancet 2013.

This is not included as it is deemed not comparable to a single intervention such as mNutrition. The relevant ranges for mNutrition are (in $ 2012):

- micronutrient interventions: $20–100/DALYs averted;
- community management of severe acute malnutrition (less relevant): $26–39/DALYs averted; and
- education programmes on nutrition/WASH: already included (Waters *et al.,* (2006)), but not with DALYs metrics: $100 /DALYs averted.

These metrics are the go-to numbers for the cost-effectiveness of interventions, so it may also be good to acknowledge the mNutrition metrics against them for comparison. mNutrition may be best equated to the education on WASH/nutrition category, given it is an advice-based service rather than a direct medical intervention service.

However, the challenge with the cost-effectiveness metrics published in the DCP is that they are far removed from the details and context of the studies from which they are derived. Several instances of this that are relevant to the metrics for nutrition-based interventions are laid out above.

For instance, the education metric is based on just one publication, Waters *et al.,* (2006). Following up on the detail reveals that the publication describes a study in Peru, consisting of a health facility-based nutrition education programme. The programme included complementary feeding demonstrations, nutrition messages, and growth monitoring sessions. This is a broader programme than mNutrition, and one that is managed from health facilities. Knowing this, it might be considered less appropriate to compare mNutrition to the education programmes on the WASH/nutrition metric in DCP3.

Secondly, the method behind the cost-effectiveness calculations, whether using a collection of studies or derived and converted from just one study, is not readily available. Following references in DCP3 to the source of the calculations, one is referred to a working paper by Horton *et al.,* 2015.

Horton *et al*., (2015) contains a list of the studies that went into each metric, in the form of a list of references of papers (this is how it is possible to track that Waters *et al*., (2006) is the reference for the education programmes intervention). Unfortunately, the table that should contain the cost-effectiveness metrics compiled from all the studies included is missing, with the tag ‘WILL FOLLOW LATER’. Likewise, the method used to aggregate metrics or convert estimates from one form to another, is not given. (Perhaps this is because the authors did simple averaging, but even in this case it should be mentioned.) On the plus side, the method used to convert results into $ 2012 is clearly explained.

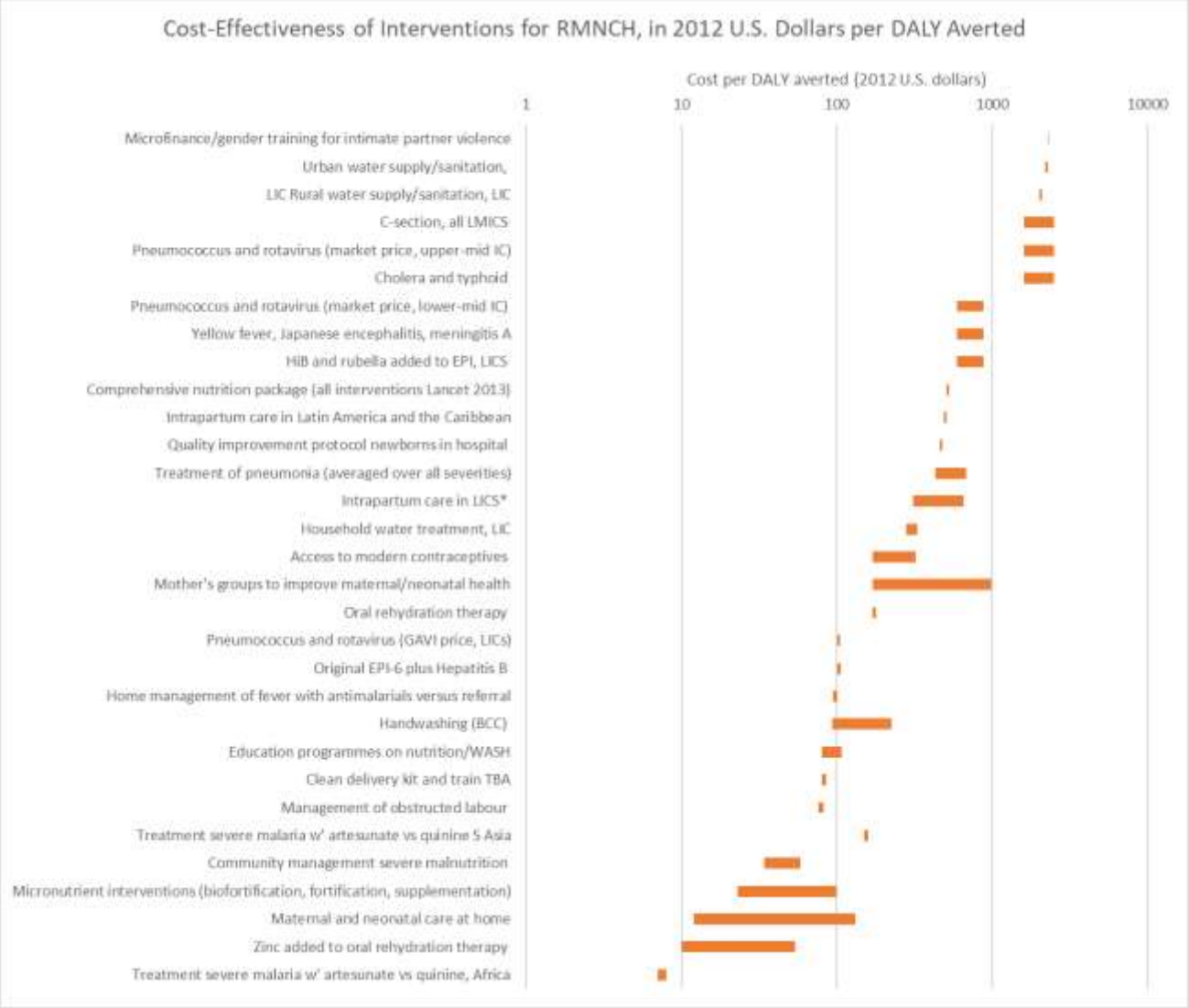
Referring back to Waters *et al*., (2006), from which the education programmes for WASH/nutrition metric is derived, the cost-effectiveness metrics are reported as cost per child, cost per stunting case averted, and cost per death averted. The conversion to DALYs is done for the DCP3, though how is not made clear. Looking to other literature by the same authors, the Copenhagen Consensus Challenge Paper by Horton *et al*., (2008) offers a clue. The paper reports the original metrics from Waters *et al*., (2006) and then converts them into DALYs: $61–153/DALYs averted, with the note that a death in Peru is equivalent to 32 DALYs. This paper is entirely unconnected to the DCP3, apart from being authored by the same person. The lack of documentation regarding method, and transparency of assumptions, is troubling.

From a methodological point of view a useful discussion about DALYs as a metric is included in Chapter 17. DCP3 refers to the DALYs measure as an ‘unresolved issue’. In 2012, the Institute of Health Metrics and Evaluation recommended that DALYs should not be discounted or an age

weighting applied. Before then, discounting and age-weighting was the norm. This means that when drawing comparisons, there is room for confusion as to how DALYs have been calculated, and whether or not comparisons are possible: 'Already, the lack of a single outcome measure makes comparisons of interventions more difficult, and this recent methodological advice will exacerbate the difficulties.' Horton 2015.

Given these limitations, nevertheless Horton (2015) offers a simple graphic from which the reader can get a feel for the spread of cost per DALYs averted.

Figure 23: Cost-effectiveness of interventions for RMNCH, in 2012 $ per DALYs averted.



Source: Authors' own. Derived from Horton *et al.,* (2015)

Note: EPI = Expanded Program of Immunization: Gavi = Global Alliance for Vaccines and Immunization: HiB = Haemophilus influenzae B; LIC = low-income country; a. Converted from life years, assuming that a newborn life is equivalent to 32 DALYs or 60 life years.

C.3 Studies from the Ghana CEA registry

Five studies from the eight found through the Ghana CEA registry are in the health field. Only one of these is relevant for sub-Saharan Africa, the rest relating to India, Pakistan, and Bangladesh. None of them are clear matches for mNutrition, being about food and micronutrient supplements, fortified food, cash transfers, and community management of malnutrition (and none of these being

situated in sub-Saharan Africa). Potentially of most relevance to mNutrition, due to the location and intervention type, is Bergmann *et al.,* (2017), which examines the integration of nutrition and HIV services in Malawi and Mozambique. In Malawi, the three interventions included the use of SMS technology: messaging to remind caregivers to attend appointments, texting between clinic staff and community members, relaying of test results to the clinic quickly, and tracking nutrition status of children. It is not clear that this extends to actual advice being given, as is the format of mNutrition, but cost-effectiveness metrics are available for the suite of three interventions together (which includes male motivators and child health passports, in addition to the SMS technology), and for each intervention individually. In terms of methodology, Bergmann *et al.,* (2018) openly acknowledge that they only pick the outcomes that are measurable. The costs used in the analysis are complex and not easy to understand because the intervention is question is a complex one – see the paper for more details.

C.4 What do people do when there is no, or minimal, significant outcome effect?

Multiple studies from the 18 added to the mHealth database from the literature survey faced challenges in the economic evaluation because no significant impact of the intervention on the chosen measured outcome(s) was apparent. The studies apply a range of solutions to this problem, in some cases enabling the authors to continue with the economic evaluation, and in some cases requiring a pivot in evaluation methodology from the original protocol.

Two studies describe how no significant health outcomes were measured, and so the economic analysis pivots to focus on other aspects of value for money.

The MAM'Out trial sought to assess the effectiveness of multi-annual and seasonal cash transfers in regard to the prevention of moderate acute malnutrition in children under three years of age (Tonguet-Papucci *et al.*, (2015)). This study was conducted in Burkina Faso and one of the key aims of the trial was to assess the cost-effectiveness of the intervention. Unfortunately, no effect on moderate acute malnutrition was found: 'We were not able to assess cost-effectiveness of the mobile transfer intervention due to there being no significant impact on the primary outcome of acute malnutrition incidence; though impacts were found on secondary outcomes… a positive impact was demonstrated on child nutrient intake, and child and caregiver dietary diversity' (Houngbe *et al.,* (2017), referencing Tongeut-Papucci *et al*., (2017) for the secondary outcomes data). Therefore, no CEA was possible according to the authors' original aims of examining stunting as the outcome. Instead, the authors assessed the cost efficiency of the mobile cash transfer intervention, given that there were positive secondary outcomes (Puett *et al*., 2018). Unlike other studies (Kasteng *et al*., (2018), discussed below), they did not change the CEA to use the secondary outcomes.

Nisbett *et al.,* (2016), in the evaluation of a FCDO programme in Bangladesh to integrate nutrition interventions into existing livelihoods programmes, discussed above, also found no effect on the health outcome (stunting) on the basis of which they planned to run a CEA. In this case, there was also a lack of impact on secondary outcomes, so the option of evaluating the cost-effectiveness in relation to other outcomes was also not possible. Instead, the authors describe how they changed their approach (with agreement from FCDO), directly referencing the other aspects of FCDO's value for money framework: efficiency, economy, equity, and (somehow) effectiveness. This means the authors did not quantify a cost-effectiveness metric but were able to assess other aspects of how the intervention was implemented, such as if resources were used well, the extent to which marginalised groups were included, and what learning and policy recommendations were possible from the findings.

Both the MAM'Out trial and Nisbett *et al*., (2016) demonstrate a change in approach to the economic evaluation, steering away from quantifying cost-effectiveness, acknowledging the lack of results for the primary outcome found, and focusing on other value for money aspects. Three other studies discussed subsequently manoeuvred their methodology or assumptions to enable a cost-effectiveness calculation regardless of insignificant effects in relation to primary outcomes.

C.5 How are DALYs interpreted?

Willcox *et al.,* (2017) is an example of a study where morbidity is ignored, and DALYs averted are calculated from mortality only. The study investigates the effectiveness of low-dose, high-frequency training on emergency obstetric and newborn care, examining newborn mortality. The authors' defence of this approach is not entirely clear. They make the obvious statement that the impact of newborn mortality is derived from years life lost, rather than years lost to disability. It is possible that they meant to say that the majority of the impact of newborn disorders is derived from years life lost, rather than years lost to disability. They also point out that neonatal disorders are the leading cause of death for under-fives, but only the fourth contributing factor of years lost to disability, and that, in general, in sub-Saharan Africa the ratio of years lost to disability to DALYs remains low. The approach of neglecting years lost to disability serves to make their analysis more conservative.

C.6 What studies were not included in the literature database?

Five studies were found that do have cost-effectiveness metrics, but that for various reasons were not included in the database. They have entries in the search notes below. One of these (Johns *et al.*, 2019) used life years saved only and did not include morbidity. Given this study was also far from being comparable to mNutrition in terms of the type of intervention (strengthening of district health services) and aim (did not address nutrition), this was not included. Okafor *et al*., (2017) looked at diarrhoea management approaches in Nigeria. Diarrhoea is a major contributing factor to malnutrition and can also result from it (see DCP3), and so it is an interesting question as to whether mNutrition should be compared to studies that are 'nutrition-sensitive', or whether this is a degree too distant in terms of aims and type of intervention. Okafor *et al*., (2017) was not included but is noted as a potential option if it becomes relevant.

The closest study to mNutrition but which is not included is Prinja *et al*., (2018) (the protocol is Prinja *et al*., (2016)). This examines ReMIND, an mHealth intervention which is a job-aid for community workers. It tracks their clients and guides the community workers through a woman's pregnancy and newborn childcare. Although this is an mHealth study, it was not included because: it is a supply-side mHealth intervention and so is not similar to mNutrition; it is based in India; and the CEA was unclear (the authors have data for five years, but apply a model for 10 years, and it is not clear why or what parameters the model uses).

Huda *et al*., (2018) assess the effectiveness of mobile phone nutrition counselling with cash transfers in a pilot study in Bangladesh. The effectiveness is limited to whether the women received and understood the messages, rather than changes to practices or health outcomes. However, in 2018 the authors started a four-year randomised control trial which will examine effectiveness in terms of preventing stunting, so this may be useful as a comparator in the future.

Ruel-Bergeron *et al*., (2019) report a further project which may be comparable to mNutrition: Malawi's Right Foods and the Right Time programme uses behaviour change communication on nutrition, supplements, and heath to improve health outcomes. The study reports a monitoring and evaluation framework, and mentions cost-effectiveness, but does not report any results at this stage. The trial finished in 2018 so these results may be forthcoming.

Projects called Tubaramure, in Burundi, and Procomida, in Guatemala, have three associated publications, one of which references that a CEA will be presented in a forthcoming publication. The interventions are food rations, behaviour change communication, and encouragement to use health services, aimed at women and children. For Burundi, Leroy *et al*., (2016) report a positive impact on the number of cases of anaemia, and Leroy *et al*., (2018) report a reduction in child stunting. For Guatemala, Olney *et al*., (2018) report a reduction in stunting, and state that a cost analysis will be published separately. There are therefore a number of trials that might be comparable for mHealth and mAgri, but which are not available currently for this evaluation.

Gimaiyo *et al*., (2019) describe a joint sanitation and nutrition programme in Kitui, Kenya, called SanNut, and so is another study, alongside Hasan *et al,.* (2018, that examines the linkage between sanitation and nutrition). SanNut ran alongside standard community-led total sanitation programmes (this is an approach used to get the community involved and active in improve sanitation outcomes). It provided additional education on how hygiene affects nutrition, good practices for children's hygiene and food practices, and other health aspects. No impacts were found on nutritional practices, and only modest improvements were found in sanitation knowledge and practices, so the integration is concluded to be unsuccessful. The control arm was communities that received only the community-led total sanitation, so it was concluded that at least SanNut did not detract from community-led total sanitation, and actually enhanced it in terms of sanitation outcomes, despite having no impact on nutrition outcomes.

C.7 References for literature update

Ali, S., Vargas, P., and Keen, S. (2017) 'Cost-effectiveness of the WINNN Programme. Operations Research and Impact Evaluation', Oxford.

Bar-Zeev, N., for the VACSURV Consortium, Tate, J. E., for the VACSURV Consortium, Pecenka, C., for the VACSURV Consortium, … Whitney, C. G. (2016) 'Cost-Effectiveness of Monovalent Rotavirus Vaccination of Infants in Malawi: A Postintroduction Analysis Using Individual Patient–Level Costing Data', *Clinical Infectious Diseases, 62*(suppl_2), pp. S220–S228. https://doi.org/10.1093/cid/civ1025

Bergmann, J. N., Legins, K., Sint, T. T., Snidal, S., Amor, Y. B., McCord, G. C., and UNICEF Research Group (2017) 'Outcomes and Cost-Effectiveness of Integrating HIV and Nutrition Service Delivery: Pilots in Malawi and Mozambique', *AIDS and Behavior 21*(3), pp. 703–711. https://doi.org/10.1007/s10461-016-1400-3

Black, R. E., and Dewey, K. G. (2019) 'Benefits of supplementation with multiple micronutrients in pregnancy', *Annals of the New York Academy of Sciences*, *1444*, pp. 3–5. https://doi.org/10.1111/nyas.14088

Bourassa, M. W., Osendarp, S. J. M., Adu-afarwuah, S., Ahmed, S., Ajello, C., Bergeron, G., … Dewey, K. G. (2019) 'Review of the evidence regarding the use of antenatal multiple micronutrient supplementation in low- and middle-income countries', *Annals of the New York Academy of Sciences*, *1444*, pp. 6–21. https://doi.org/10.1111/nyas.14121

Bowser, D. M., Shepard, D. S., Nandakumar, A., Okunogbe, A., Morrill, T., Halasa-rappel, Y., … Erhunmwunse, O. A. (2018) 'Cost Effectiveness of Mobile Health for Antenatal Care and Facility Births in Nigeria', *Annals of Global Health*, *84*(4), pp. 592–602.

DFID (2011) 'DFID's Approach to Value for Money (VfM)', DFID.

Engle-stone, R., Kumordzie, S. M., Meinzen-dick, L., and Vosti, S. A. (2019) 'Replacing iron-folic acid with multiple micronutrient supplements among pregnant women in Bangladesh and Burkina Faso: costs, impacts, and cost-effectiveness', *Annals of the New York Academy of Sciences*, *1444*, pp. 35–51. https://doi.org/10.1111/nyas.14132

Fiedler, J. L., Lividini, K., Drummond, E., and Thilsted, S. H. (2016) 'Strengthening the contribution of aquaculture to food and nutrition security: The potential of a vitamin A-rich, small fish in Bangladesh', *Aquaculture 452*, 291–303. https://doi.org/10.1016/j.aquaculture.2015.11.004

Freeman, P. A., Schleiff, M., Sacks, E., Rassekh, B. M., Gupta, S., and Perry, H. B. (2017) 'Comprehensive review of the evidence regarding the effectiveness of community – based primary health care in improving maternal, neonatal and child health: 4. child health findings', *Journal of Global Health*, *7*(1). https://doi.org/10.7189/jogh.07.010904

Garg, C. C., Mazumder, S., Taneja, S., Shekhar, M., Mohan, S. B., Bose, A., … Bhandari, N. (2018) 'Costing of three feeding regimens for home-based management of children with uncomplicated severe acute malnutrition from a randomised trial in India', *BMJ Global Health 3*(2). https://doi.org/10.1136/bmjgh-2017-000702

Gimaiyo, G., Mcmanus, J., Yarri, M., Singh, S., Trevett, A., Moloney, G., … Lehmann, L. (2019) 'Can child-focused sanitation and nutrition programming improve health practices and outcomes? Evidence from a randomised controlled trial in Kitui County, Kenya', *BMJ Global Health*, *4*, pp. 1–12. https://doi.org/10.1136/bmjgh-2018-000973

Goudet S., Jayaraman A., Chanani S., Osrin D., Devleesschauwer B., Bogin B., … Griffiths P., (2018) 'Cost effectiveness of a community based prevention and treatment of acute malnutrition programme in Mumbai slums, India'. *PloS one*, *13*(11) 1–17. https://doi.org/10.1371/journal.pone.0205688

Green, T., Michaux, K., Moumin, N., and Talukder, Z. (2018) 'Final Technical Report: Scale up of Homestead Food Production for Improved Household Food Security and Nutrition in Cambodia - Fish on Farms Phase 2: Family Farms For The Future'.

Grimes, C. E., Law, R., Dare, A., Day, N., Reshamwalla, S., Murowa, M., … Lavy, C. B. (2017) 'Cost-Effectiveness of Two Government District Hospitals in Sub-Saharan Africa', *World Journal of Surgery*, *41*(9), pp. 2187–2192. https://doi.org/10.1007/s00268-017-4007-6

Hagen-Zanker, J., Bastagli, F., Harman, L., Barca, V., Sturge, G., and Schmidt, T. (2016) 'Understanding the impact of cash transfers: the evidence'. Overseas Development Institute. Retrieved from www.odi.org/sites/odi.org.uk/files/resource-documents/10748.pdf

Haghparast-Bidgoli, H., Skordis, J., Harris-Fry, H., Krishnan, S., O'Hearn, M., Kumar, A., … Ojha, A. K. (2019) 'Protocol for the cost-consequence and equity impact analyses of a cluster randomised controlled trial comparing three variants of a nutrition-sensitive agricultural extension intervention to improve maternal and child dietary diversity and nutritional status in rural Odisha, India (UPAVAN trial).', *Trials*, *20*, pp. 1–11.

Hansford, F., Visram, A., Jones, E., and Ward, P. (2017) 'Integrated Evaluation Report of the WINNN Programme: Operations Research and Impact Evaluation', Oxford, UK.

Harman, L., Bastagli, F., Hagen-zanker, J., Sturge, G., and Barca, V. (2016) 'Cash transfers: what does the evidence say? An annotated bibliography'.

Horton, S., Alderman, H., Rivera, J. A., Horton, S., Alderman, H., and Rivera, J. A. (2008) 'Copenhagen Consensus 2008 Challenge Paper, Hunger and Malnutrition' (pp. 1–40). pp. 1–40. Copenhagen Consensus Center.

Horton, S., Wu, D. C. N., Brouwer, E., & Levin, C. (2015). Methods and Results for Systematic Search, Cost, and Cost-Effectiveness, Disease Control Priorities, third edition. Working Paper Series

Huda, T. M., Alam, A., Tahsina, T., Hasan, M. M., Khan, J., Rahman, M. M., … Dibley, M. J. (2018) 'Mobile-based nutrition counseling and unconditional cash transfers for improving maternal and child nutrition in Bangladesh: Pilot study', *JMIR mHealth and uHealth*, *6*(7), p.e156. https://doi.org/10.2196/mhealth.8832

IFPRI (2014) '2014 Global Nutrition Report Data'. Retrieved from: http://ebrary.ifpri.org/utils/getfile/collection/p15738coll2/id/128484/filename/128695.pdf%0Ahttp://dx.doi.org/10.7910/DVN/27857

Isanaka, S., Barnhart, D. A., McDonald, C. M., Ackatia-armah, R. S., Kupka, R., Doumbia, S., … Menzies, N. A. (2019) 'Cost-effectiveness of community-based screening and treatment of moderate acute malnutrition in Mali', *BMJ Global Health 4*, pp. 1–10. https://doi.org/10.1136/bmjgh-2018-001227

Johns, B., Hangoma, P., Atuyambe, L., Faye, S., Tumwine, M., Zulu, C., … Mugasha, C. (2019) 'The Costs and Cost-Effectiveness of a District-Strengthening Strategy to Mitigate the 3 Delays to Quality Maternal Health', *Global Health: Science and Practice 7*, pp. 104–122.

Kashi, B., M Godin, C., Kurzawa, Z. A., Verney, A. M. J., Busch-Hallen, J. F., and De-Regil, L. M. (2019) 'Multiple Micronutrient Supplements Are More Cost-effective than Iron and Folic Acid: Modeling Results from 3 High-Burden Asian Countries', *The Journal of Nutrition.* https://doi.org/10.1093/jn/nxz052

Kasteng, F., Murray, J., Cousens, S., Sarrassat, S., Steel, J., Meda, N., … Borghi, J. (2018) 'Cost-effectiveness and economies of scale of a mass radio campaign to promote household life-saving practices in Burkina Faso', *BMJ Global Health 3*, pp. 1–11. https://doi.org/10.1136/bmjgh-2018-000809

King, J. (2017) 'Using Economic Methods Evaluatively', *American Journal of Evaluation 38*(1), pp. 101–113. https://doi.org/10.1177/1098214016641211

King, J., and Allan, S. (2018) 'Applying evaluative thinking to value for money: The Pakistan Sub-national Governance Programme', *Evaluation Matters—He Take Tō Te Aromatawai* 4, p. 1. https://doi.org/10.18296/em.0034

King, J., and Guimaraes, L. (2016) 'Evaluating value for money in international development: the Ligada female economic empowerment programme in Mozambique', *eVALUation Matters*, 3, pp. 58–69.

King, J., and OPM. (2018) 'The OPM approach to assessing Value for Money: A guide', OPM, Oxford.

Krämer, M., Kumar, S., and Vollmer, S. (2018) 'Impact of delivering iron-fortified salt through a school feeding program on child health, education and cognition: Evidence from a randomized controlled trial in rural India', *GlobalFood Discussion Paper No. 116*.

Krämer, M., Kumar, S., and Vollmer, S. (2018) 'Improving Children Health and Cognition: Evidence from School-Based Nutrition Intervention in India', *Discussion Paper No. 247*, Courant Research Centre: Poverty, Equity and Growth, Gottingen.

Leech, A. A., Kim, D. D., Cohen, J. T., and Neumann, P. J. (2018) 'Use and Misuse of Cost-Effectiveness Analysis Thresholds in Low- and Middle-Income Countries: Trends in Cost-per-DALY Studies', *Value in Health* 21(7), pp. 759–761. https://doi.org/10.1016/j.jval.2017.12.016

LeFevre, A. E., Mohan, D., Hutchful, D., Jennings, L., Mehl, G., Labrique, A., … Moorthy, A. (2017) 'Mobile Technology for Community Health in Ghana: what happens when technical functionality threatens the effectiveness of digital health programs?' *BMC Medical Informatics and Decision Making 17*(27), pp. 1–17. https://doi.org/10.1186/s12911-017-0421-9

LeFevre, A., Cabrera-Escobar, M. A., Mohan, D., Eriksen, J., Rogers, D., Parsons, A. N., … Coleman, J. (2018) 'Forecasting the Value for Money of Mobile Maternal Health Information Messages on Improving Utilization of Maternal and Child Health Services in Gauteng, South Africa: Cost-Effectiveness Analysis', *JMIR Mhealth Uhealth 6*(7), pp. 1–12. https://doi.org/10.2196/mhealth.8185

Masset, E., Mascagni, G., Acharya, A., Egger, E.-M., and Saha, A. (2018) 'The Cost-Effectiveness of Complex Projects: A Systematic Review of Methodologies', *IDS Bulletin* 49(4). https://doi.org/10.19088/1968-2018.166

Moss, C., Bekele, T. H., Salasibew, M. M., Sturgess, J., Ayana, G., Kuche, D., … Dangour, A. D. (2018) 'Sustainable Undernutrition Reduction in Ethiopia (SURE) evaluation study: a protocol to evaluate impact, process and context of a large-scale integrated health and agriculture

programme to improve complementary feeding in Ethiopia', *BMJ Open 8*, pp. 1–11. https://doi.org/10.1136/bmjopen-2018-022028

Nair, N., Tripathy, P., Sachdev, H. S., Pradhan, H., Bhattacharyya, S., Gope, R., … Prost, A. (2017) 'Effect of participatory women's groups and counselling through home visits on children's linear growth in rural eastern India (CARING trial): a cluster-randomised controlled trial', *The Lancet Global Health 5*(10). https://doi.org/10.1016/S2214-109X(17)30339-X

Nisbett, N., Longhurst, R., Barnett, I., Feruglio, F., Gordon, J., Hoddinott, J., … White, J. (2016) 'Impact Evaluation of the DFID Programme To Accelerate Improved Nutrition for the Extreme Poor in Bangladesh: Final Report'.

Okafor, C. E., and Ekwunife, O. I. (2017) 'Cost-effectiveness analysis of diarrhoea management approaches in Nigeria: A decision analytical model', *PLoS Neglected Tropical Diseases 11*(12), pp. 1–14.

OPM and King, J. (2018) 'In depth: OPM's approach to assessing Value for Money', OPM, Oxford, UK. https://doi.org/10.1161/CIRCULATIONAHA.116.021889

OPM (2017) 'Bihar Child Support Programme (BCSP) Endline Impact Evaluation Report', OPM, Oxford, UK. Retrieved from www.opml.co.uk/sites/default/files/BCSP-Endline-Report.pdf.

OPM (2017) 'Child Development Grant Programme: Key messages of the midline evaluation findings (Summary Report)', OPM, Oxford, UK.

Patel, A., Kuhite, P., Puranik, A., Khan, S. S., Borkar, J., and Dhande, L. (2018) 'Effectiveness of weekly cell phone counselling calls and daily text messages to improve breastfeeding indicators', *BMC Pediatrics 18*(1), pp. 1–12. https://doi.org/10.1186/s12887-018-1308-3

Plessow, R., Arora, N. K., Brunner, B., and Wieser, S. (2016) 'Cost-effectiveness of price subsidies on fortified packaged infant cereals in reducing iron deficiency anemia in 6-23-month-old-children in Urban India', *PLoS ONE 11*(4), pp. 1–21.https://doi.org/10.1371/journal.pone.0152800

Prinja, S., Bahuguna, P., Gupta, A., Nimesh, R., Gupta, M., and Thakur, J. S. (2018) 'Cost effectiveness of mHealth intervention by community health workers for reducing maternal and newborn mortality in rural Uttar Pradesh, India', *Cost Effectiveness and Resource Allocation* 16(25), pp. 1–19. https://doi.org/10.1186/s12962-018-0110-2

Rogers, E., Martínez, K., Luis, J., Morán, A., Alé, F. G. B., Charle, P., … Puett, C. (2018) 'Cost-effectiveness of the treatment of uncomplicated severe acute malnutrition by community health workers compared to treatment provided at an outpatient facility in rural Mali', *Human Resources for Health 16*(12), pp. 1–9.

Rothstein, J. D., Jennings, L., Moorthy, A., Yang, F., Gee, L., Romano, K., … LeFevre, A. E. (2016) 'Qualitative Assessment of the Feasibility, Usability, and Acceptability of a Mobile Client Data App for Community-Based Maternal, Neonatal, and Child Care in Rural Ghana', *International Journal of Telemedicine and Applications*.

Ruel-Bergeron, J. C., Hurley, K. M., Kang, Y., Aburto, N., Farhikhtah, A., Dinucci, A., … Christian, P. (2019) 'Monitoring and evaluation design of Malawi's Right Foods at the Right Time nutrition program', *Evaluation and Program Planning*, *73* (November 2018), pp. 1–9. https://doi.org/10.1016/j.evalprogplan.2018.11.001

Russell, L. B., Kim, S., Cosgriff, B., Ram, S., Schrag, S. J., Meulen, A. S., … Sinha, A. (2017). 'Cost-effectiveness of maternal GBS immunization in low-income sub-Saharan Africa', *Vaccine* 35, pp. 6905–6914. https://doi.org/10.1016/j.vaccine.2017.07.108

Schreinemachers, P., Patalagsa, M. A., and Uddin, N. (2016) 'Impact and cost-effectiveness of women's training in home gardening and nutrition in Bangladesh', *Journal of Development Effectiveness* 8(4), pp. 473–488. https://doi.org/10.1080/19439342.2016.1231704

Sinha, R. K., Haghparast-Bidgoli, H., Tripathy, P. K., Nair, N., Gope, R., Rath, S., and Prost, A. (2017) 'Economic evaluation of participatory learning and action with women's groups facilitated by Accredited Social Health Activists to improve birth outcomes in rural eastern

India', *Cost Effectiveness and Resource Allocation 15*(1), pp. 1–9. https://doi.org/10.1186/s12962-017-0064-9

Skordis-worrall, J., Sinha, R., Ojha, A. K., Sarangi, S., Nair, N., Tripathy, P., … Haghparast-bidgoli, H. (2016) 'Protocol for the economic evaluation of a community-based intervention to improve growth among children under two in rural India (CARING trial)', *BMJ Open* 6. https://doi.org/10.1136/bmjopen-2016-012046

Svefors, P., Selling, K. E., Shaheen, R., Khan, A. I., Persson, L. Å., and Lindholm, L. (2018) 'Cost-effectiveness of prenatal food and micronutrient interventions on under-five mortality and stunting: Analysis of data from the MINIMat randomized trial, Bangladesh', *PLoS ONE* 13(2), pp. 1–15. https://doi.org/10.1371/journal.pone.0191260

The Kenya CT-OVC Evaluation Team (2012) 'The impact of Kenya's Cash Transfer for Orphans and Vulnerable Children on human capital', *Journal of Development Effectiveness 4*(1), pp. 38–49. https://doi.org/10.1080/19439342.2011.653578

Trenouth, L., Colbourn, T., Fenn, B., Pietzsch, S., Myatt, M., and Puett, C. (2018) 'The cost of preventing undernutrition: Cost, cost-efficiency and cost-effectiveness of three cash-based interventions on nutrition outcomes in Dadu, Pakistan', *Health Policy and Planning 33*(6), pp. 743–754. https://doi.org/10.1093/heapol/czy045

Walters, D. D. (2018) 'Economics of Maternal and Child Nutrition: Food Fortification with Vitamin A in Tanzania', (Doctoral dissertation), University of Toronto.

Walters, D., Ndau, E., Mosha, T., and Horton, S. (2019) 'Cost-effectiveness of sunflower oil fortification with vitamin A in Tanzania by scale', *Maternal & Child Nutrition 15*(S3), pp. 1–11. https://doi.org/10.1111/mcn.12720

Wang, Y. H., Zou, C. Q., Mirza, Z., Li, H., Zhang, Z. Z., Li, D. P., … Zhang, Y. Q. (2016) 'Cost of agronomic biofortification of wheat with zinc in China', *Agronomy for Sustainable Development* 36(3). https://doi.org/10.1007/s13593-016-0382-x

Waters, H. R., Penny, M. E., Creed-Kanashiro, H. M., Robert, R. C., Narro, R., Willis, J., … Black, R. E. (2006) 'The cost-effectiveness of a child nutrition education programme in Peru', *Health Policy and Planning*, *21*(4), pp. 257–264. https://doi.org/10.1093/heapol/czl010

Wieser, S., Brunner, B., Tzogiou, C., Plessow, R., Zimmermann, M. B., Farebrother, J., … Bhutta, Z. A. (2018) 'Reducing micronutrient deficiencies in Pakistani children: Are subsidies on fortified complementary foods cost-effective?' *Public Health Nutrition*, *21*(15), 2893-2906.

Willcox, M., Harrison, H., Asiedu, A., Nelson, A., Gomez, P., and Lefevre, A. (2017) 'Incremental cost and cost-effectiveness of low-dose, high-frequency training in basic emergency obstetric and newborn care as compared to status quo: part of a cluster- randomized training intervention evaluation in Ghana', *Globalization and Health 13*(88), pp. 1–14. https://doi.org/10.1186/s12992-017-0313-x

Willcox, M., Moorthy, A., Mohan, D., Romano, K., Hutchful, D., Mehl, G., ... LeFevre, A. (2019) 'Mobile Technology for Community Health in Ghana: Is Maternal Messaging and Provider Use of Technology Cost-Effective in Improving Maternal and Child Health Outcomes at Scale?' *Journal of Medical Internet Research 21*(2), pp. 1–13. https://doi.org/10.2196/11268

Woods, B. (2016) 'Woods *et al.* (2016) Supplementary Information'.

Woods, B., Revill, P., Sculpher, M., and Claxton, K. (2016) 'Country-Level Cost-Effectiveness Thresholds: Initial Estimates and the Need for Further Research', *Value in Health 19*(8), pp. 929–935. https://doi.org/10.1016/j.jval.2016.02.017

Xie, R.-H., Tan, H., Taljaard, M., Liao, Y., Krewski, D., Du, Q., and Wen, S. W. (2018) 'The Impact of a Maternal Education Program Through Text Messaging in Rural China: Cluster Randomized Controlled Trial', *JMIR MHealth and UHealth 6*(12), p. e11213. https://doi.org/10.2196/11213

Zhang, C. M., Zhao, W. Y., Gao, A. X., Su, T. T., Wang, Y. K., Zhang, Y. Q., … He, X. H. (2018) 'How Could Agronomic Biofortification of Rice Be an Alternative Strategy with Higher Cost-

Effectiveness for Human Iron and Zinc Deficiency in China?' *Food and Nutrition Bulletin 39*(2), pp. 246–259. https://doi.org/10.1177/0379572117745661