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

## Mobile phones, nutrition, and health in Tanzania: Quantitative endline report

Daniel O. Gilligan, Melissa Hidrobo, Giordano Palloni, and Heleene Tambet
(International Food Policy Research Institute)

14 October 2020

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

## Acknowledgements

The International Food Policy Research Institute (IFPRI) thanks all the households in Iringa, Tanzania, that agreed to participate in this research project. We are also grateful to the Oxford Policy Management Tanzania (OPMT) team for leading the data collection, and to Groupe Spéciale Mobile 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 support for the quantitative component of the overall impact evaluation.

The IFPRI research team benefitted enormously from our ongoing partnership with the Institute of Development Studies (IDS), which oversees the qualitative component of the evaluation, and the Gamos team conducting the business model portion of the evaluation. Both teams have provided invaluable input and feedback as we have designed the evaluation, developed tools, and composed this report. Finally, we would also like to thank the Foreign, Commonwealth and Development Office (FCDO) and Oxford Policy Management (OPM) teams for their continued support and contributions to the project.





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 Daniel O. Gilligan, Melissa Hidrobo, Giordano Palloni, and Heleene Tambet. 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 | Tel | +44 (0) 1865 207300 |
| | 52 Cornmarket Street | Fax | +44 (0) 1865 207301 |
| | Oxford OX1 3HJ | Email | admin@opml.co.uk |
| | United Kingdom | 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 Open and Enhanced 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 from FCDO, the endline report will be made available on IFPRI and IDS's websites. All datasets will also be made available within 12 months of final data collection on IFPRI's page at the Harvard Dataverse website: https://dataverse.harvard.edu/dataverse/IFPRI.

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 Daniel O. Gilligan, Melissa Hidrobo, Giordano Palloni, and Heleene Tambet. 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 | Tel | +44 (0) 1865 207300 |
| | 52 Cornmarket Street | Fax | +44 (0) 1865 207301 |
| | Oxford OX1 3HJ | Email | admin@opml.co.uk |
| | United Kingdom | Website | www.opml.co.uk |

# Executive summary

## The mNutrition service and evaluation design

### mNutrition in Tanzania

The mNutrition initiative is a five-year global programme supported by the UK Foreign, Commonwealth and Development Office (FCDO), organised and supported by Groupe Spéciale Mobile Association (GSMA), and implemented by in-country mobile network operators (MNOs) and third-party providers, which aims to use mobile technology to improve the health and nutritional status of children and adults in low-income countries around the world. mNutrition 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 that are likely to result in improved nutritional health within a household.

The mNutrition intervention that is the focus of the evaluation in Tanzania, and of this report, is an integrated service that combines an existing short message service (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 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 was 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 of Development Studies (IDS), and the International Food Policy Research Institute (IFPRI). The team draws on a variety of different tools and methods to collect evidence on the impact of the mNutrition intervention in Tanzania. Broadly speaking, 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. The evaluation is being conducted in the Iringa Region of Tanzania.

This report focuses on the quantitative component, which employed a cluster randomised controlled trial (cRCT) 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 during a promotion campaign coordinated by the study team; households in villages randomly assigned to the control group did not receive any offer of access to the service through the study's promotion campaign. However, households in either treatment group could have learned about the programme through other means, including through other promotional activities run by the HPHB team and its partners.

The quantitative evaluation design included other components to provide evidence on factors affecting take-up and use of the service. In addition to the village-level randomisation, the evaluation included a second-stage household-level randomisation among eligible households in treatment villages to help us to better understand how nutrition information flows between spouses. For households in treatment villages where both the pregnant woman or the mother of the child under 12 months (the primary female) and her spouse or another adult male decision-maker in the household (the primary male) owned distinct mobile phones at baseline, we randomly assigned half of the households to receive the mNutrition service only on the phone of the primary female household member, and the other half of the households to receive the mNutrition service on both the phone of the primary female household member and the phone of the primary male. Comparing outcomes between these two sub-treatment groups and the control households that would have been eligible for the sub-randomisation allows us to explore whether sending the information to multiple household members differentially affects household beliefs and behaviours related to nutrition.

In practice, we estimate 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 from the study's promotion campaign—known as 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, we calculate the local average treatment effect (LATE) on compliers (households that were induced to receive the mNutrition content by the random door-to-door offer), which under additional assumptions approximates the impact of receiving the mNutrition messages on outcomes for 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. We are reluctant to rely too heavily on the self-reported measures of service exposure and therefore interpret the LATE estimates as upper bounds of the true LATE impacts for compliers.

The quantitative evaluation answers the following research questions:

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 nutritional status for children under 12 months at baseline?
4. What is the impact of the mNutrition service on nutrition knowledge, including knowledge of IYCF practices, among pregnant women and the caregivers of young children?
5. Does sending the mNutrition content to the mobile phones of both the primary female—either the pregnant woman or the mother of the child under 12 months at baseline—and the primary male—typically the spouse of the primary female—have a differential impact on the other primary and secondary outcomes?

To measure the quantitative impacts of the programme on the primary and secondary outcomes, baseline and endline surveys were conducted. Oxford Policy Management Tanzania (OPMT) served as the in-country survey partner, leading the data collection in cooperation with the quantitative evaluation team from IFPRI. The baseline survey took place between October and

December 2016 and collected information from 2,833 households across 180 villages. The endline survey targeted the same set of households and successfully interviewed 2,595 between October 2018 and January 2019. A two-year gap between the baseline and endline surveys was selected to ensure that beneficiary households would be exposed to the mNutrition content for long enough that updated beliefs and behaviours could potentially affect the nutrition outcomes of children. Child anthropometry, in particular, is likely to require sustained behaviour change throughout the first 1,000 days of children's lives (de Onis et al., 2013).

## Service exposure and experiences with mNutrition

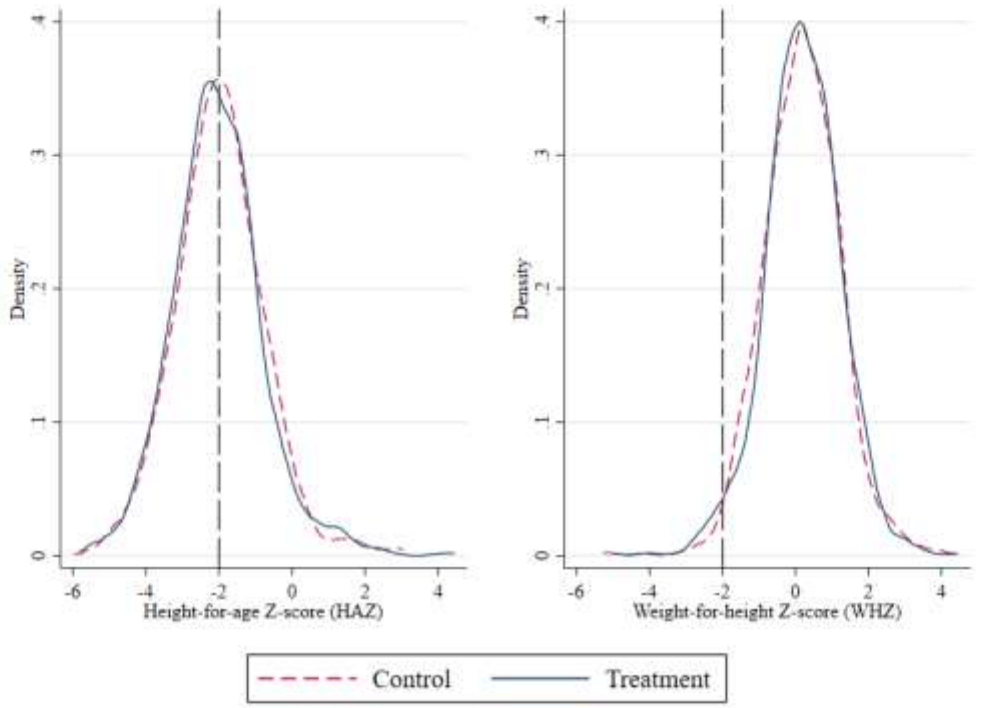
Relative to control households, **households in villages randomly assigned to the mNutrition treatment group were 39.7 percentage points more likely to report having received text messages with nutrition information** (66.4% compared to 26.7%) and 29.9 percentage points more likely to identify the messages as being from the mNutrition service (42.7% compared to 12.8%) during the study period. Administrative data on the registration rate of mobile phone numbers for treatment households early in the study period indicate that the true gap in exposure between the two groups may be considerably larger than the implied gap based on self-reports, and may potentially approach a difference of 70 percentage points.

Individuals in treatment households that reported receiving the mNutrition content indicated that they received the messages with varying frequency: 38.3% of primary females and 46.1% of primary males reported receiving mNutrition messages less than once per month, while 33.9% of primary females and 28.0% of primary males reported receiving messages at least once per week during the study period.

**mNutrition participants were likely to read the messages they received and reported a high level of satisfaction with the message content.** 85.8% of primary females and 86.1% of primary males reported reading all the mNutrition messages that they received on their mobile phones, with another 9.7% of females and 9.5% of males indicating that they read at least some of the content they were sent. Perceptions of the content of the mNutrition messages were nearly universally positive: 91.4% of females and 93.2% of males found the messages either always useful or very often useful.

## The impact of mNutrition on primary and secondary outcomes

**Figure 1.1: Endline child anthropometry by mNutrition treatment group**




Source: Authors' own

## ITT estimates

We find **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, and wasting): the ITT estimates clearly indicate zero impacts. Furthermore, as shown in Figure 1.1, the distributions for HAZ and WHZ are nearly indistinguishable across the treatment groups at endline. Results for the other primary and secondary outcomes are more positive, however. **mNutrition improved dietary diversity and adequacy for children in communities randomly assigned to receive an offer of access to the mNutrition service**. Children aged 6–35 months in treatment villages consume from 0.107 more food groups and are 3.8 percentage points more likely to satisfy the minimum dietary diversity (MDD) threshold. Children aged 6–23 months from treatment villages are also 6.9 percentage points more likely to achieve a minimum acceptable diet (MAD). Similarly, **women's diets in treatment households are somewhat more diverse**. While the impact on the number of food groups consumed is not statistically significantly different from zero, treatment women are 4.0 percentage points more likely to meet the minimum dietary diversity for women (MDD-W) threshold, suggesting that they are more likely to have nutritionally adequate diets. We observe these dietary improvements for women and children despite there being no evidence of changes in household-level food consumption patterns.

**mNutrition improved male nutrition knowledge in treatment households.** We estimate a positive and statistically insignificant effect on female knowledge and a statistically significant positive treatment effect for males: treatment group males answer 1.7% more IYCF and nutrition knowledge assessment questions correctly than control group males. The effect size is equal to 3.5% of the overall baseline mean, indicating that male knowledge improvements, while statistically significant, were not large in magnitude. A combined household-level IYCF knowledge score is also statistically significantly increased by mNutrition access, with treatment households scoring

0.09 standard deviations higher on the endline knowledge assessment, equivalent to answering 1.1% more questions correctly, on average.

The main ITT evidence therefore suggests that the mNutrition messaging impacted beliefs and knowledge and modestly improved 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 and dietary diversity—on their own—were not large enough to affect child nutritional status. We caution that while we intentionally limited the number of outcomes, inference was not adjusted within families of outcomes for multiple hypothesis testing.

### LATE estimates

The LATE estimates measure the impact of mNutrition on the subset of participants that were induced to join the service because of the offer of access from the study's promotional campaign (compliers). The LATE results show that *receipt* of the nutrition messages during the previous two years has 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 messages had on these same outcomes (as captured by the ITT estimates).[1] Despite the LATE estimates being larger in magnitude than the ITT estimates, the LATE estimates similarly suggest that receipt of the mNutrition messaging did not have any significant impacts on children's nutritional status.

While both the ITT and LATE estimates are potentially useful for policymakers, we prefer the ITT estimates for several reasons. The most relevant is that the measures of programme participation—whether households self-report having received nutrition-related text messages or mNutrition content during the study period—are imperfect measures of true receipt of the mNutrition content. If the measurement error in relation to programme participation results in an underestimate of the true difference in mNutrition exposure between treatment and control households, the LATE estimates will be an upper bound of the true LATE. As the treatment–control gap in mNutrition message receipt approaches one, the difference between the ITT and the LATE messages shrinks to zero.

## The impact of mNutrition on information sources, trust, and mobile phone use

Beyond the primary and secondary outcomes, we also use the cRCT design to identify the impact of the offer of access to the mNutrition service on mobile phone use and expenditure, sources of nutrition information, and trust in nutrition information received through different mediums. **For both females and males, exposure to the mNutrition service increased mobile phone use.** We find meaningful impacts on the likelihood of sending and receiving text messages (for both males and females), and making or receiving phone calls (for females), and an increased likelihood of receiving mobile money (for females), as well as an increased likelihood of sending mobile money (for males).

Consistent with the increased use of mobile phones, treatment households also report spending more on mobile phone airtime vouchers. The impact estimates indicate that **the offer of access to**

[1] With the offer of access to the mNutrition service increasing reported receipt of nutrition-related text messages by 35–40 percentage points, the LATE estimates for complier households are between 2.5 and 2.85 times as large as the ITT estimates as, by construction, the LATE estimates will equal the ratio of the ITT estimates to the difference in service take-up between treatment and control households.

**the service increased mobile phone spending by just under 10% during the month preceding the endline survey**. Using conservative assumptions about how mobile phone vouchers are allocated by households across available mobile networks, this implies there was a large increase in mobile phone spending on the main mobile network as a result of exposure to the mNutrition service.

We find robust evidence that males and females in treatment households are 17.9 percentage points (males) and 23.0 percentage points (females) more likely to report getting any nutrition information through text messages, as well as being 10.0 percentages points (males) and 10.4 percentages points (females) more likely to identify automatic text messages as one of their primary sources of nutrition information.

## Policy Implications

The rapid expansion of access to mobile phones across the developing world is an exciting development for policymakers. Information gaps—low levels of knowledge or the high prevalence of inaccurate information—can potentially be addressed in a low-cost manner using mobile phones as a platform to disseminate knowledge. However, while appealing in theory, there exist barriers that may limit the usefulness of mobile phones in affecting knowledge, changing behaviour, and improving welfare for individuals in developing countries. The limited positive impacts discussed here complimented by the more in-depth discussion of potential barriers like low mobile phone attachment, poor network service, low levels of trust in non-human sources, low message salience, and unequal access to mobile phones in Barnett et al. (2020), suggest there are real constraints on the effectiveness of mobile phones for improving health and nutrition outcomes.

Nevertheless, that the stand-alone SMS-based information service studied here was able to produce some observable changes in knowledge and behaviour highlights that mobile phones can be an important tool for development practitioners. Understanding how the mNutrition service could be adapted to improve nutrition outcomes (e.g. HAZ and stunting) among children in beneficiary households is an important avenue for future research. Pairing the mobile messaging with interventions meant to overcome other barriers to proper nutrition seems like a promising path forward. For example, SMS-based content could be combined with in-person BCC meetings that target practices like breastfeeding that may be difficult to change solely through the provision of basic information. Alternatively, if household resource constraints are a critical determinant of child undernutrition (which existing research indicates is the case), the service could be linked to resource transfers of cash or food in order to help households maximise the potential nutrition benefits of the transfer.

**Key highlights**

- Men in treatment villages answered 1.7% more nutrition knowledge questions correctly than men in control villages, while women in treatment villages did not answer statistically significantly more nutrition knowledge questions correctly than their control group counterparts.
- Access to the mNutrition content improves dietary diversity for children aged 6–35 months by 0.11 (food groups, out of 10) and increases the likelihood that these children satisfy MDD by 3.8 percentage points.
- For children aged 6–23 months, the offer of access to the mNutrition content increases the likelihood of meeting the MAD threshold by 6.9 percentage points.
- Dietary diversity for women also increases due to the mNutrition service: the primary females in treatment households are 4.0 percentage points more likely to satisfy the MDD-W standards than primary females in control group households.
- Despite the changes in knowledge and behaviours, the random offer of access to the mNutrition service had no impact on the anthropometry of sample children: we estimate small and statistically insignificant impacts on HAZ, WHZ, stunting, and wasting. This suggests that the main barriers to early childhood growth failure in rural Iringa may not be information-related, or that the observed changes in nutrition knowledge alone were not large enough to improve child nutritional status.
- We find no evidence that sending the mNutrition content to the mobile phone of the primary male, in addition to the mobile phone of the primary female, has any additional impact on the other primary or secondary outcomes.
- Females and males in treatment villages make greater use of basic mobile services because of exposure to the mNutrition service; on average, households spend Tanzanian shillings (TSH) 510.5 (9.6%) more per month on mobile phone airtime vouchers.

# Table of contents

# List of figures and tables

# List of abbreviations

| | |
|---|---|
| ANCOVA | Analysis of Covariance |
| BCC | Behaviour Change Communication |
| CAPI | Computer-Assisted Personal Interviewing |
| CDC | US Centers for Disease Control and Prevention |
| CDDS | Child Dietary Diversity Score |
| CLE | Community Listing Exercise |
| COSTECH | Commission for Science and Technology |
| COUNSENUTH | Centre for Counselling, Nutrition and Health Care |
| cRCT | Cluster Randomised Controlled Trial |
| CTB | Convex Time Budget |
| DFID | UK Department for International Development |
| DHS | Demographic and Health Survey |
| FCDO | UK Foreign, Commonwealth and Development Office |
| GAIN | Global Alliance for Improved Nutrition |
| GSMA | Groupe Spéciale Mobile Association |
| HAZ | Height-for-Age Z-Score |
| HDDS | Household Dietary Diversity Score |
| HPHB | Healthy Pregnancy, Healthy Baby |
| ICT | Information and Communication Technology |
| IDS | Institute of Development Studies |
| IFPRI | International Food Policy Research Institute |
| IRB | Institutional Review Board |
| ITT | Intent-to-Treat |
| IYCF | Infant and Young Child Feeding |
| JHCCP | Johns Hopkins Center for Communication Programs |
| LATE | Local Average Treatment Effect |
| MAD | Minimum Acceptable Diet |
| MDD | Minimum Dietary Diversity |

| | |
|---|---|
| MDD-W | Minimum Dietary Diversity 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 |
| NGO | Non-Governmental Organisation |
| NPL | National Poverty Line |
| OLS | Ordinary Least Squares |
| OPM | Oxford Policy Management |
| OPMT | Oxford Policy Management Tanzania |
| PPI | Poverty Probability Index |
| RCT | Randomised Controlled Trial |
| SIM | Subscriber Identity Module |
| SMS | Short Message Service |
| TFNC | Tanzania Food and Nutrition Centre |
| THBS | Tanzania Household Budget Survey |
| TSH | Tanzanian Shillings |
| WAZ | Weight-for-Age Z-Score |
| WDDS | Women's Dietary Diversity Score |
| WHZ | Weight-for-Height Z-Score |
| 2SLS | Two-Stage Least Squares |

# 1 Introduction

A mounting body of evidence links early childhood undernutrition to increased morbidity and mortality (Pelletier *et al.*, 1995), as well as to poor adult outcomes, including shorter stature, decreased educational attainment, reduced economic productivity (Alderman *et al.*, 2006; Victora *et al.*, 2008; Hoddinott *et al.*, 2013), and increased incidence of non-communicable disease (Barker *et al.*, 1989; Gluckman and Hanson, 2004). Despite the potentially serious consequences, early childhood malnutrition remains common around the world: as of 2011, 165 million children under the age of five were stunted and 52 million children under the age of five were wasted (Black *et al.*, 2013).

Although the causes are inarguably complex, poor maternal nutrition during pregnancy (Black *et al.*, 2013; Christian *et al.*, 2013) and inadequate IYCF practices (Bhutta *et al.*, 2008) are thought to be two of the principal drivers of early childhood undernutrition. Improving these behaviours therefore seems likely to generate important returns for both childhood nutrition and adult well-being.

With the rapid increase in access to and ownership of mobile phones across sub-Saharan Africa and the broader developing world (Poushter *et al.*, 2015), information and communication technology (ICT) interventions using mobile phones are increasingly seen as a feasible way to disperse information to individuals and households. Largely, though not exclusively, these campaigns have focused on improving farmers' information for agriculture through the provision of crop and input prices, weather information, and agricultural extension services (Svensson and Yanagizawa, 2009; Fafchamps and Minten, 2012; Courtois and Subervie, 2014; Hildebrandt *et al.*, 2015; Aker *et al.*, 2016; Cole and Fernando, 2012).

Though less common, ICT interventions, and specifically SMS-based information interventions, have also been used to provide health-related information (Labrique *et al.*, 2013). Typically, these interventions target improved patient drug adherence (Nglazi *et al.,* 2013) or behaviour change related to sexual and reproductive health (Rokicki *et al.*, 2017). Few SMS-based message campaigns have targeted nutrition-related behaviour change. Jiang *et al.* (2013) and Flax *et al.* (2014) tested whether two such interventions influenced IYCF practices in Nigeria and China, respectively. To date, the existing research on ICT for nutrition and health finds mixed results on effectiveness, and the nutrition-focused studies have not been designed to test for ICT-related impacts on child nutrition outcomes.

mNutrition, a global initiative supported by FCDO, organised by GSMA, and implemented by in-country service management organisations in cooperation with MNOs, explores the potential to use mobile technology to change attitudes, knowledge, behaviours, and practices for improved nutritional status. In Tanzania, the service 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 is intended 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 are being conducted by a consortium of researchers from Gamos, IDS, and IFPRI. The team draws 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 cRCT 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), which aims to provide an understanding of mNutrition's context, its underlying mechanisms of change, and its implementation process.
- 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 estimate the wider imputed benefits from the value-added service for the range of stakeholders involved.

The quantitative component of the evaluation is designed to contribute evidence to help answer the first two broad questions specified in the Terms of Reference (see Annex A), while the full mixed-methods evaluation design will address all six of the research questions in the Terms of Reference (TOR):

1. 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?

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?

Individual technical reports have been written to address each component – quantitative, qualitative, and business model and cost effectiveness – with each addressing different parts of the TOR. The quantitative component of the evaluation is best suited to address the first two study questions. The quantitative evaluation's approach to generating rigorous impact estimates of the programme's effects using a causal research design will directly respond to question 1. These impact estimates will be combined with information on the cost of implementing the programme components to construct estimates of cost-effectiveness. The surveys from the quantitative evaluation will also provide the information needed to answer question 2 on the effectiveness of the service at reaching target groups and changing their knowledge and behaviours around nutrition.

There have been no major divergences from the research questions specified in the TOR, though some additional questions have been added to the quantitative component (see Section 1.2). A mixed-methods report combines the findings of the three evaluation components to build a deeper understanding, gather lessons learned, and inform best practices about the future design and implementation of mobile phone-based information services to ensure (a) behaviour change and (b) continued private sector engagement in different countries (Barnett et al., 2020).

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 therefore comprises pregnant women and caregivers of children under the age of five years who reside in rural areas of the study region (Iringa).

For the sample of pregnant women and caregivers of young children in rural Iringa selected to participate in the quantitative study, the evaluation focuses on estimating the causal effect of access to the mNutrition service. That is, the evaluation will identify how nutrition-related behaviours, knowledge, and outcomes are altered for service beneficiary households relative to their counterfactual levels, i.e. what the value of the outcome would have been for beneficiary households in the absence of access to the mNutrition service.

The intended audience for the quantitative endline report is FCDO, along with other organisations involved in mNutrition and mHealth programmes globally (including local MNOs and non-governmental organisations (NGOs) implementing mNutrition services), national governments—in particular, the MoHCDGEC and the TFNC in Tanzania—international agencies and donors, and community-level health workers. Consultations and workshops with stakeholders occurred throughout the evaluation period, with in-country visits prior to the start of the evaluation to help finalize the design, and a workshop after the completion of the baseline data collection to present key baseline findings. The findings of this report were presented and discussed with key stakeholders during a webinar in May 2020. The reports from the evaluation, which have been reviewed and commented on by stakeholders, will be publicly available on IFPRI and IDS's websites, and will continue to be disseminated through webinars, international conferences, journal publications, and blog posts.

## 1.2 Research questions

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 will be 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 will be addressed through the quantitative evaluation are:

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 nutritional status for children under 12 months at baseline?

In addition, the evaluation will also address the following secondary research questions:

4. What is the impact of the mNutrition service on nutrition knowledge, including knowledge of IYCF practices, among pregnant women and the caregivers of young children?

5. Does sending the mNutrition content to the mobile phones of both the primary female—either the pregnant woman or the mother of the child under 12 months at baseline—and the primary male—typically the spouse of the primary female—have a differential impact on the other primary and secondary outcomes?

The three primary research questions specify the main outcomes that will be 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). The first secondary research question (Question 4) addresses overall study Question 2 on the effectiveness of the interventions for improving nutrition knowledge and behaviours. The next secondary research question (Question 5) addresses both of the overall study questions by examining how changing the target recipient of information in the household affects the impact and cost-effectiveness of the programme across each of the study outcomes.

The remainder of this document proceeds as follows. Section 2 discusses the mNutrition service being evaluated and provides basic information about the study context. Section 3 presents the evaluation design and empirical strategy. Section 4 summarises the endline data collection. Section 5 presents statistics on self-reported interaction with the mNutrition service overall and by treatment group. Section 6 presents the main ITT impact estimates on the primary and secondary outcomes. Section 7 presents the LATE of the mNutrition service on households induced to participate in the service by the random offer of access. Section 8 presents the results based on the household-level sub-randomisation and explores whether there is any heterogeneity in the impact estimates by baseline household wealth. Section 9 presents ITT estimates of mNutrition on mobile phone use, sources of nutrition information, and trust in nutrition information received from different sources. Finally, Section 10 concludes.

# 2 mNutrition intervention

The mNutrition initiative is a five-year global programme supported by FCDO, organised and supported by the GSMA, and implemented by in-country MNOs and third-party organisations. It aims 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 that are likely to result in improved nutritional health within a household.

In Tanzania, the mNutrition service that is the focus of the evaluation and of this report is an integrated programme that enhances HPHB, an existing SMS-based health communication campaign targeting pregnant women and mothers of young children. The mass media programme accompanying the service is called Wazazi Nipendeni. Wazazi Nipendeni is a US Center for Disease Control and Prevention- (CDC-) funded project bringing together multiple partners contributing toward shared goals. Phase 1 of mNutrition, launched in 2012, was initially developed in coordination with the Tanzania Capacity Communication Project, a United States Agency for International Development-funded service led by the Johns Hopkins Center for Communication Programs (JHCCP). 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 JHCCP, while the SMS component of the campaign was led by the mHealth Tanzania Public–Private Partnership (mHealth Tanzania-PPP). The mHealth Tanzania-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 sends free text messages containing healthcare information to pregnant women, mothers with newborns, and male supporters and general information seekers in Tanzania to drive health-seeking behaviour (Open Government Partnership, n.d.). 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, which allow 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 that are delivered to mothers or caregivers of children up to five years old. At the beginning of 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 will be referred to as the mNutrition service in the remaining sections of this report.

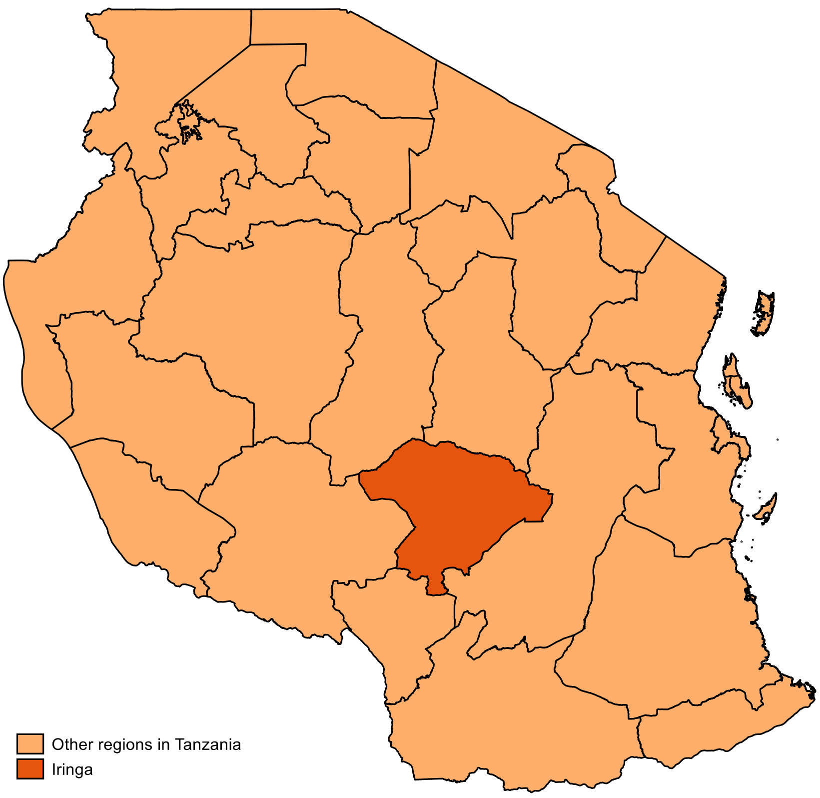
The 120 nutrition messages included in the mNutrition service are drawn from 42 factsheets on nutrition-related behaviours identified as key determinants of outcomes that were developed by GAIN together with local partners the Centre for Counselling, Nutrition and Health Care (COUNSENUTH) and Every1mobile. The information contained in these factsheets was adapted to the context of Tanzania and made mobile-ready by the local content providers under the guidance of the MoHCDGEC and TFNC. As 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 providing new information for the target populations 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-sourced 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 being tobacco, cashew nuts, coffee, cotton, and sesame seed (Food and Agriculture Organization of the United Nations, 2018). Tanzania is divided into 31 regions and these 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 aged between 6 and 59 months are anaemic (DHS, 2016).

The evaluation was carried out within the three rural districts of Iringa region in Tanzania: Iringa rural, Kilolo, and Mufindi. Figure 2.1 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. In 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 Iringa region: roughly 89% of households in Iringa rural, 92% of households in Mufindi, and 92% of households in Kilolo are involved in agriculture. The average household size is 4.2 and the adult literacy rate is 79% among the rural population (Population and Housing Census, 2012). As in Tanzania more broadly, child undernutrition is a severe problem in Iringa: according to the 2016 DHS, 41.6% of children under the age of five were stunted. This figure was nearly seven 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 aged 6–59 months were anaemic.

**Figure 2.1: Iringa region in Tanzania**




Source: Global Administrative Areas (2015). GADM database of Global Administrative Areas, version 2.8. [online] URL: www.gadm.org

In part, Iringa was selected as the location for the study because of a dearth of existing relationships between 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). Although 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 (see Gilligan *et al.*, 2018).

# 3 Evaluation design

## 3.1 Study design

This section draws heavily on Sections 3.1 and 3.2 of the baseline report, which provides detailed information about each step in the design of the quantitative evaluation (Gilligan *et al.*, 2018).

The quantitative evaluation was designed as a cRCT, with two stages of randomisation: a village-level randomisation where villages are assigned to a treatment group or to a control group, and a household-level randomisation within treatment villages whereby households are 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, and there was some Tanzania-wide advertising of the service through billboards and radio adverts, 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.

Randomised controlled trials (RCTs) are widely viewed as the most reliable method of quantitative programme evaluation. By allowing researchers to randomly select who is affected by or offered access to a programme, RCTs provide a clean way to estimate the causal effect of the programme. In comparison, observational studies are forced to compare outcomes across individuals or groups that *choose* to participate in the programme to individuals or groups that do not. In most circumstances, this simple difference will be the sum of the causal effect of participating in the programme and a term representing how the individuals or groups that choose to participate are different from those that do not, even in the absence of exposure to the programme. This latter term, typically referred to as selection bias, can be positive or negative depending on the context and the intervention being studied. Selection bias is eliminated by the random assignment of treatment (or a treatment offer) in an RCT: the individuals or groups included in the study are allocated to receive the treatment (or treatment offer) by the randomisation mechanism controlled by the research team. On average, the units assigned to the treatment group and the units assigned to the control group should therefore have the same outcome levels, in the absence of exposure to the treatment, as well as similar levels of other characteristics that could affect the outcomes. As a result, any difference in outcomes at the end of an RCT can be attributed to the intervention being evaluated and not to selection bias.

To ensure that the evaluation accurately measures the causal impact of access to the mNutrition service, the quantitative evaluation was based on a cRCT. 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 includes 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 both 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, learning can be generated about how information flows between spouses.

Prior to baseline fieldwork, sample size calculations were undertaken to identify the number of study clusters to include in the evaluation. To facilitate the identification and sampling of potential beneficiary households in both treatment and control villages, baseline data collection included two separate activities. First, a community listing exercise (CLE) was conducted to gather basic information to determine whether households were eligible to be sampled. Next, a baseline household survey collected detailed information on all outcomes, as well as on other characteristics likely to be predictive of outcomes for sampled households. Between the CLE and household survey, the research team processed the CLE data to identify eligible households, randomly sampled households from those deemed eligible, and conducted the village-level randomisation in such a way as to minimise the expected variance of the treatment effects.

The quantitative evaluation included two rounds of household surveys: a baseline survey completed between October and December 2016 and an endline survey with the same set of households completed between October 2018 and January 2019. A two-year gap was chosen to allow for a sufficient exposure to the mNutrition service. Generating impacts on child anthropometry, for instance, is likely to require sustained behaviour change throughout the first 1,000 days of children's lives (De Onis *et al.*, 2013). Both rounds of data collection served a distinct purpose. The main functions of the baseline household survey were to assess balance in observable characteristics across treatment groups, to collect baseline measures of outcomes to be used in endline analyses in order to improve statistical power through Analysis of Covariance (ANCOVA) or difference-in-differences specifications, to collect baseline measures of non-outcome characteristics to reduce residual variance and potentially improve the precision of the treatment effect estimates at endline, and, finally, to explore the relevance of the mNutrition service for the study population. The main objectives of the endline household survey, on the other hand, were to measure primary and secondary outcomes in order to estimate the causal impacts of the service, and to connect treatment effects to levels and changes in key individual and household characteristics to better understand causal pathways.

## 3.2 Overview of sample design

### 3.2.1 Household eligibility

In Tanzania, the mNutrition messages have been integrated within the existing HPHB programme, which targeted general health information to pregnant women, mothers of children up to age five, and their caregivers. Although the mNutrition content is designed to benefit children up to the age of two, prior to the evaluation we elected to restrict potential beneficiary households to those with either a pregnant woman or a child aged under 12 months[2] at the time of the CLE. This age restriction was made to ensure that all of the focus children included in the sample would have at least one full year of exposure to the service during their first 1,000 days of life (Gilligan *et al*., 2018). A subset of the mNutrition content is targeted toward pregnant women—including SMS reminders to take iron supplements, consume iodised salt, and remain active during pregnancy—and other messages focus on behaviours that need to occur immediately following childbirth; for

[2] The mother of the child under 12 months was also required to be a household member.

example, the service includes messages encouraging the mother to initiate breastfeeding as soon as possible after the birth. For these reasons, we set two distinct sampling targets in each study cluster: one for households with pregnant women (six per cluster) and one for households with a child under the age of 12 months (11 per cluster).

In addition to eligibility restrictions related to the age of the focus child or the pregnancy status of the primary female, several other eligibility criteria were used to identify the study sample. First, to be eligible, households were required to have at least one household member that was literate in Swahili. Second, at least one member of the household had to own a mobile phone. These restrictions were made to ensure that the mNutrition content—which is sent to mobile phones via SMS messages in Swahili—would be accessible to everyone included in the sample. In practice, these additional restrictions were rarely binding: of the 4,689 households identified as having either a pregnant woman or the mother of a child under the age of 12 months, 94.7% also owned at least one mobile phone and had a member who was literate in Swahili. All households that met the three eligibility criteria were retained in the potential study sample from which the final sample was later selected.

### 3.2.2 Village sampling

The quantitative evaluation grouped together all households residing within the same village by designating each sampled village as a distinct study cluster and ensuring that the treatment status was the same for all households in the same cluster. The resulting cRCT provides less statistical power but smaller risk of spill-overs (i.e. one unit's treatment assignment having a direct effect on another unit's outcomes) than a RCT that assigns treatment status at the individual level. In determining the level at which to cluster observations, three potential consequences were considered. First, the content of mNutrition SMS messages is relatively easy and costless to share within a community. Since this type of information-sharing spill-over across individuals within the same village is not problematic only in a case where all households in the same village are assigned to the same treatment, we ensured that treatment allocation was constant within villages through a cluster randomised design. Note that this does not completely preclude the possibility of across district spill-overs through information-sharing between households in different villages, but it does reduce the likelihood that this will occur. Second, it was necessary to identify a sufficient number of eligible households in each study cluster. Piloting the baseline questionnaire as well as available data from the 2010 DHS confirmed that most villages in Iringa would likely contain enough eligible households to meet these per cluster targets. Third, we had to address the trade-off between broad study clusters—that help to reduce spill-overs and increase the likelihood of having a sufficient number of eligible households—and loss in statistical power for a given fieldwork budget. Ultimately, the budget for the quantitative evaluation was sufficient to cover fieldwork costs within one region of Tanzania, and there were too few available divisions, districts, wards, and health clinics within the regions being considered for us to use any of these to define study clusters. More details on identifying the most suitable clustering structure can be found in Gilligan *et al.*, (2018).

### 3.2.3 Sample size

Power calculations were conducted to determine the sample size necessary to identify meaningfully sized treatment effects of the mNutrition service on HAZ for children. The evaluation is designed to detect impacts on HAZ because HAZ for young children has been identified as both a good indicator of undernutrition and a meaningful predictor of health and well-being later in life (Black *et al.*, 2013). In addition, HAZ measurement relies solely on the data collection team, not on accurate reporting by a household member. Therefore, relative to self-, mother-, or household member-reported measures of health and nutrition, HAZ is less vulnerable to problems related to measurement error or reporting bias.

To determine the optimal sample size, critical moments of the HAZ distribution and the intra-cluster correlation were estimated using data from the 2010 DHS in Tanzania, and autocorrelation was set based on calculations from the NEEP-IE study in Malawi (Gelli *et al.*, 2017) and the PROMIS study in Mali (Huybregts *et al.*, 2017). Reasonable assumptions were made about the household attrition rate; application of these assumptions generated a take-up gap between treatment and control villages; ultimately, we determined that for the evaluation to be able to detect differences in HAZ between the treatment and control groups of at least 0.25 standard deviations it was necessary that at least 180 village clusters be included in the study. The exact descriptions of the full set of parameters assumed or calculated for power calculations can be found in the baseline report (Gilligan *et al*., 2018).

To identify the villages to include in the quantitative evaluation, our data collection partner OPMT provided us with a list of all rural villages in Iringa region of Tanzania. Rural areas with emerging urban characteristics were dropped from the list, which left us with 354 rural villages that were eligible to be selected for the study sample. Of these 354 eligible villages we randomly selected 180 to be included in the main sample and an additional 20 villages to be used as potential replacements. Ultimately, two of the villages originally selected for the main sample were replaced due to there being no mobile network coverage anywhere in these two villages.

## 3.3 Estimation strategy[3]

### 3.3.1 ITT estimates

Because the offer of treatment was randomly assigned, we use the systematic variation in the random offer of access to the mNutrition service 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 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—we use ANCOVA to generate our primary estimates, and we use simple differences and difference-in-differences models as robustness checks.

For outcomes that were not observable at baseline—such as children's dietary diversity, for example—we use a simple differences specification to estimate treatment effects. To identify the simple differences treatment effects, we estimate the following ordinary least squares (OLS) regressions:

$$Y_{1ihv} = \beta_0 + \beta_{1,SD} Treatment_v + \delta_x X_{0ihv} + \varepsilon_{ihv} \,,$$

where $Y_{1ihv}$ is the outcome measured at endline ($t = 1$), for individual $i$, in household $h$, in village $v$, $\beta_0$ is a constant term, $Treatment_v$ is an indicator equal to 1 if village $v$ was randomly assigned to the treatment group, $X_{0ihv}$ is a vector of observable characteristics for individual $i$ measured at baseline ($t = 0$), which includes the baseline measure of the outcome when available, and $\varepsilon_{ihv}$ is

[3] The following section draws heavily from the mNutrition Tanzania baseline report (Gilligan *et al.*, 2018) and is discussed in more detail in Annex D.

an error term that we cluster at the village level. In this model, $\delta_x$ represents the vector of coefficients on the controls and $\beta_{1,SD}$ is the parameter of interest: the causal effect of being offered access to the mNutrition service on the outcome.

With one baseline survey and one endline survey, treatment effects estimated through an ANCOVA specification—that is, that include the baseline measure of the outcome as a control variable–are expected to have lower variance[4] than the difference-in-differences treatment effects, as long as the autocorrelation for the outcome is below 1.[5]

### 3.3.2 LATE for compliers

The specifications described above enable us 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[6] we can also estimate the LATE of receiving the mNutrition content for compliers, i.e. 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.[7] Both of these assumptions are likely to be satisfied in the context of this evaluation (see the full discussion in Gilligan *et al*., (2018)).

If both assumptions are satisfied and participation in the mNutrition service is reliably observed in the data, the LATE for complier households can be estimated through two-stage least squares (2SLS) estimates of the ANCOVA or simple difference models discussed in the previous subsection, using the random assignment of villages to receive the mNutrition treatment offer as an excluded instrument for observed take-up of the service.[8]

The LATE treatment effects for compliers provide a different, but still policy-relevant parameter: they represent, albeit for a specific sub-population (compliers), the causal effect of exposure to the mNutrition messaging. Regardless, in the following sections we estimate and discuss both parameters in order to provide more complete conclusions about the causal effects of the mNutrition service.

[4] Although we know the variance of the treatment effects estimated through ANCOVA will be lower than the variance for those estimated with simple differences or difference-in-differences, we cannot know the size of the difference between the variances until we have collected endline data and estimated the autocorrelations for each outcome.
[5] See McKenzie (2012). In an experiment with one baseline and one follow-up survey, the ratio of the variance of the difference-in-differences estimate to the variance of the ANCOVA variance is given by $\frac{2}{(1+\rho)}$, where ρ is the autocorrelation.
[6] See Imbens and Rubin (2015) for a complete discussion.
[7] 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 to register for the service without having received the offer.
[8] For the linear models specified in this context, the 2SLS estimates of LATE for compliers will be equal to the ratio of the ITT estimate of the impact of the treatment offer on the outcome to the ITT estimate of the impact of the treatment offer on take-up of the programme.

## 3.4 Overview of baseline balance

Baseline data were successfully collected from 2,833 households, or 92.6% of the 3,060 originally targeted households. The random allocation of villages to the treatment or control group, and treatment households to sub-treatment groups, ensures that a simple difference between treatment arms yields unbiased estimates of the causal effect of the service. However, in addition to the clear theoretical basis for estimating causal impacts through a randomised experiment, we also used the data collected through the baseline survey to empirically assess how likely it is that the random assignment to treatment arms would be successful in identifying similar groups of households. As imbalance in observable attributes at baseline, especially any imbalance thought to be strongly correlated with the outcomes of interest, typically casts doubt on the ability of the evaluation to identify the causal effect of the intervention being investigated, we compared baseline measures of the primary and secondary outcomes—as well as other observable characteristics likely to be correlated with the outcomes—across treatment groups. We determined two sets of balance measures for each baseline characteristic: i) the p-value from a t-test of a null hypothesis that there is no difference in means between the two treatment groups; and ii) the normalised difference (the ratio of the difference in means to the average of the within treatment group standard deviation) suggested by Imbens (2015). Following Imbens' approach, we interpreted normalised differences below 0.25 as being indicative of baseline balance.

The baseline balance analysis of sample characteristics for household demographics, physical structure of the home, amenities, and household wealth revealed remarkable balance in characteristics across the two treatment arms. Of the 26 characteristics tested, just one difference was significant at the 5% level. Even more reassuring, the normalised differences were extremely small in magnitude: none were above the 0.25 threshold and only two out of 26 had a normalised difference above 0.10. The two treatment groups were also well balanced in terms of baseline measures of the anthropometry outcomes. The largest observed normalised difference was -0.103 (for the *Child is wasted* indicator), well below the 0.25 cut-off. Similarly, only one of the 12 tests of the null of no difference between the mean value in treatment villages and in control villages had a p-value below 0.05 (for wasting prevalence); this, again, is roughly what we would expect to observe by chance. In addition, Kolmogorov-Smirnov tests of the equality of distributions for the three main anthropometry indicators (HAZ, weight-for-age Z-scores (WAZ), and WHZ) suggested that there were no important differences in the distributions of anthropometry across the two treatment groups.

Similarly, IYCF practices at baseline were well balanced across the two treatment arms. Only one normalised difference was above the 0.25 threshold—the normalised difference for children aged 12–15 months still being fed breast milk was -0.37, suggesting that children in treatment villages were less likely to still be fed breast milk between those ages. However, there were only 102 children in total who fell within this age range and the p-value from the t-test of equality of means between the treatment and control groups fails to reject the null hypothesis at the 5% level. Only one difference was significant at the 5% level: infants were slightly more likely to be exclusively breastfed for the first six months in treatment villages (p-value 0.042). However, given the number of tests that we conducted, it was likely that there would be some imbalance eventually, and it was also reassuring that these two differences did not favour one treatment group over the other. The remainder of the normalised differences for IYCF indicators were well below the 0.25 cut-off, and it therefore seems unlikely that these differences were a symptom of more problematic differences in unobservable characteristics between the two groups. Finally, the treatment groups were also well balanced in regard to the Women's Dietary Diversity Score (WDDS).

In terms of secondary outcomes, balance in indicators of the primary female's and primary male's IYCF knowledge and beliefs was assessed. It appeared that the randomisation had indeed balanced baseline levels of IYCF knowledge across the two treatment groups, with the largest normalised difference well below 0.10, and neither of the differences in summary measures of knowledge being significant at the 5% level. Based on the balance in household demographic characteristics and in primary and secondary outcomes at baseline, we concluded that the randomisation was successful in selecting observably similar households. As we found few differences between households in the different groups, this sent a strong signal that the evaluation was likely to be able to estimate the unbiased causal impacts of the service. Full characterisation of the observable attributes of households at baseline, as well as a complete discussion of baseline balance assessment, can be found in the baseline report (Gilligan *et al*., 2018).

## 3.5 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, we condense the individual responses into indexes, which we call Anderson indices. For IYCF knowledge, we 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, we use a method proposed by Anderson (2008) to combine the female and male scores for each household. Specifically, we de-mean both scores using the female and male sample average scores, respectively, and then scale the de-meaned values by the standard deviation of the relevant score in the control group. We then calculate the weighted average of the two normalised scores, using the inverse of the covariance matrix for the normalised scores to generate the weights for each score. Relative to taking a simple average of the normalised scores, this method has the advantage of placing more weight on the score with the most independent variation. To ensure that the household measure adds information for all households, we only use households with non-missing scores for both the primary male and primary female for this portion of the analysis.

We use the same method to construct two IYCF practice index scores: one for children aged 6–23 months and one that includes children aged 6–35 months. 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 aged 0–5 months are exclusively breastfed, whether children aged 12–15 months are still being breastfed, whether children aged 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 met 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. We 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 0s 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.

## 3.6 Limitations

There are a number of limitations and caveats that are important to mention when discussing the results of the quantitative impact evaluation. While the careful design of the evaluation ensures that the results presented in this report are causal estimates of the impact of the offer of access to the mNutrition service in Iringa on the primary and secondary outcomes, there is no way to project how these results would change if the study had been conducted in a different region, a different country, or even outside of sub-Saharan Africa. The constraints to knowledge, compliance with IYCF practice guidelines, and proper child nutrition almost certainly differ widely across diverse contexts.

Similarly, the mobile phone landscape is rapidly changing across the world. How households access and interact with their mobile phones is therefore constantly in flux, and the mNutrition service impacts could differ greatly with this variation. For example, although we might expect the increased availability of smartphones to increase attachment to mobile phones, and therefore make consistently reaching households by phone more feasible, it could also reduce the use of text messages in favour of phone applications that require internet access. Though content of the type included in the mNutrition service could be adapted and pushed through smartphone apps without too much difficulty, there would remain a constant need to adjust the information to fit the most prevalent means of phone-based communication.

Another potential caveat to the design of the evaluation is that we rely heavily on self-reported outcomes, with the enumerator-verified outcomes related to child anthropometry being exceptions. For self-reported outcomes, it is possible that treatment households may simply report what they think the study team wants to hear, although being aware of better IYCF and nutrition practices would itself be an indication that the service affected IYCF knowledge among beneficiaries.

# 4 Endline data collection

The mNutrition endline household survey data collection occurred in late 2018 in Iringa region of Tanzania. Fieldwork consisted of three main elements: a household interview, anthropometric measurements for children, and an experimental Convex Time Budget (CTB) exercise that was added for households in 50 randomly sampled villages.

OPMT served as the in-country survey partner leading the endline data collection in cooperation with the quantitative evaluation team from IFPRI. OPMT has substantial experience with large-scale household surveys in Tanzania, a thorough understanding of local contextual issues, and expertise and familiarity in translating technical questionnaires into Swahili.

## 4.1 Survey instrument

When paired with the baseline data collection in the context of the randomised evaluation design, the endline instrument enables us to carefully test for causal effects of the mNutrition service at different levels of the causal chain. The endline household questionnaire was adapted from the baseline survey instrument, which the IFPRI team designed based on the initial exploratory qualitative study (Barnett *et al*., 2017), the landscaping review (Barnett *et al.*, 2016), and past experiences conducting quantitative evaluations of nutrition interventions in sub-Saharan Africa. Wherever possible, the household survey collected information from three members per household: the primary female, the primary male, and a child under three years.[9] While every household in the sample had a primary female—either the pregnant woman or the mother of the child under 12 months—at baseline, at endline a primary caregiver was interviewed instead of them in a small number of households. If interviewed, the primary caregiver was not asked questions on fertility preferences, antenatal care for past pregnancies, knowledge and beliefs about IYCF practices, HIV/AIDS awareness, mobile phone use, likelihood of trust in information from different sources, or the questions related to mNutrition service exposure.

The outline for the endline questionnaire can be found in Annex E. The survey elicited information on the following categories: geographic coordinates and basic household identifiers; household composition and demographic characteristics of household members; household assets and physical characteristics of the household structure; general health and vaccination history for children under the age of five; HIV/AIDS awareness of the primary female respondent; marriage and fertility history of the primary female respondent; desired future fertility for the primary female and the primary male respondents; mobile phone access and usage of the primary female and primary female respondents; antenatal and prenatal care for children under the age of three; knowledge and beliefs about IYCF for the primary female and primary male; IYCF practices for the two youngest children under the age of five; women's dietary diversity for the primary female; primary female trust in nutrition and health information from different sources; and direct measures of height and weight for the child under three years. The list of survey modules and respective respondents can be seen in Table 4.1.

There were several notable differences between the baseline and endline surveys. First, anthropometry measurements were recorded only for the index child (age of less than 12 months) in the baseline survey, while in the endline survey the index child and any younger siblings were measured. The endline survey instrument also included an extensive module on exposure to the

[9] Anthropometry measurements were recorded for all children who were either in utero or under the age of 12 months at the time of the baseline survey (the focus child), or the younger sibling of the focus child (under three years at the time of the endline survey). Anthropometry measurement was always contingent on receiving consent from the child's mother or father. No additional information was collected from these children.

Wazazi Nipendeni programme, administered to both the primary female and primary male when available. To ensure the data were collected for the same set of children, at endline health and vaccination questions were asked for all children under the age of seven (instead of five as at baseline), while the survey instrument had questions on current pregnancy as well as all pregnancies since 2015 (at baseline since 2013). Seven extra IYCF knowledge questions were added for both females and males to get a more complete picture of nutrition knowledge for these household members. New survey modules on monthly non-food and food expenditure were also added to the endline survey.[10]

The endline questionnaire was pre-tested in Iringa region by five members of the OPMT team as well as one member of the IFPRI evaluation team. The pre-test occurred between 5 and 9 September 2018. Throughout the process, as well as after completing the pre-test, the study team made changes to the survey items to make them more suitable for the study context, to adjust translations, to test the Computer-Assisted Personal Interviewing (CAPI) program, to ensure the question and answer option language was clear and understandable, and to finalise important study protocols, particularly around tracking of individuals and households.

Table 4.1 below lists the endline questionnaire modules, along with the intended respondent for each module.

**Table 4.1: Endline questionnaire modules**

| Module | Respondent |
| --- | --- |
| Module A: Household identification | Household head or next most responsible |
| Module B: Household composition and education | Household head or next most responsible |
| Module C: Housing and assets | Household head or next most responsible |
| Module D: General health | Primary female respondent |
| Module E: Marriage and fertility history | Part 1: Primary female respondent<br>Part 2: Primary male respondent |
| Module F: Mobile phone access and usage | Part 1: Primary female respondent<br>Part 2: Primary male respondent |
| Module G: Antenatal and postnatal care | Primary female respondent |
| Module H: IYCF knowledge and beliefs | Part 1: Primary female respondent<br>Part 2: Primary male respondent |
| Module I: Nutrition practices | Primary female respondent |
| Module J: Trust likelihood of nutrition and health information | Part 1: Primary female respondent<br>Part 2: Primary male respondent |
| Module K: Anthropometry | Anthropometry specialist |
| Module L: Interaction with the programme | Part 1: Primary female respondent<br>Part 2: Primary male respondent |
| Module M: Non-food expenditure | Primary male, if male not available then primary female |
| Module N: Food consumption and expenditure | Primary female, if female not available then person most responsible for cooking |

Source: Authors' own

[10] The previously mentioned CTB module was also added in order to estimate time and risk preference parameters for the primary females in a randomly selected sub-sample of 50 villages. We generally do not discuss the results of the CTB exercise in this report.

## 4.2 Ethics approval

IFPRI received approval from the IFPRI Institutional Review Board (IRB) for the Tanzania quantitative evaluation design described in Section 3. An application for the qualitative and quantitative surveys was submitted to and approved by the Commission for Science and Technology (COSTECH) in Tanzania and a research permit was granted. Continuations to cover the endline survey data collection and analyses were obtained from the IFPRI IRB and COSTECH on 16 October 2018 and 8 August 2018 (see Annex B to view the documentation).

As an overall guiding principle, the research team sought to conduct itself in a professional and ethical manner throughout the initial exploratory study, with respect for integrity, honesty, confidentiality, voluntary participation, impartiality, and the avoidance of personal risk. These principles were guided by the Organisation for Economic Co-operation and Development (2010) Development Assistance Committee's 'Quality Standards for Development Evaluation' and FCDO's (2011) 'Ethics Principles for Research and Evaluation', which were followed throughout the evaluation.

Informed written or oral consent was collected from all participants prior to the start of the interviews. The entire field team was trained on ethical data collection and signed an ethical conduct form prior to the start of the fieldwork. Participants received two bars of soap as compensation for their participation in the interviews.

The confidentiality of the data has been protected by recording survey interview responses using CAPI, so no hard copy versions of survey questionnaires are available. All files containing raw and analysed data are securely stored in password-protected databases. Access to the complete data is restricted to the IFPRI/IDS/Gamos evaluation team. A unique household ID was assigned to each household. The name and geographic location of the respondent was kept in a separate data file, to which only the research team had access. Anonymised versions of the data sets that exclude these personal identifiers will be made available for public access at the conclusion of the study.

## 4.3 Fieldwork and household tracking

### 4.3.1 Enumerator training and pre-testing

Enumerator training for the quantitative survey data collection occurred between 24 September and 5 October 2018 and was led by the OPMT survey team. The training included five and a half days of classroom-based training and practice in Dar es Salaam and five days of piloting in Iringa region. A separate, three-day training for anthropometric specialists was also conducted, which focused on how to properly weigh and measure children under 36 months.

The survey training was intended to provide enumerators with a complete understanding of the survey instrument and field protocols. OPMT attempted to make the training as practical as possible, to give participants frequent opportunities to practise, familiarise themselves with the instruments, and discuss questions and challenges.

A total of 40 enumerators attended the household survey training: 30 household survey enumerators, six supervisors, three reserves, and one person to administer CTB pay-outs. Most had participated in the mNutrition baseline data collection, and therefore already had an understanding of the survey instrument and the study context.

### 4.3.2 Fieldwork

The endline fieldwork occurred between 6 October and 14 December 2018. Additional visits specific to the CTB exercise continued through to the end of February 2019, but these did not involve data collection. The endline survey was timed to match the timing of the baseline survey as closely as possible, to ensure that seasonally-driven changes in household behaviour would not drive differences in measurement between the baseline and the endline survey.

The data collection team attempted to contact all 2,833 households that were successfully interviewed at baseline. In each household, enumerators sought permission to measure the height/length and weight of the focus child and any younger siblings of the focus child, and to interview the primary female and primary male.

### 4.3.3 Attrition

Attrition of households and individuals within the survey sample is a very important problem. Of the planned sample of 2,833 households across 180 villages interviewed at baseline, the data collection team eventually conducted 2,477 fully complete interviews and a further 118 that were partially completed or had a missing male interview. This implies an overall attrition rate of roughly 9%. A total of 2,803 anthropometric measurements were taken, due to the measurement of the younger siblings of index children. We present additional details on the endline survey completion and attrition, overall and by treatment group, in Table 4.2 below.

**Table 4.2: Summary of household surveys in the treatment and control group**

| | Treatment | Control | Total |
|---|---|---|---|
| Number of enumeration areas 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: Authors' own

At baseline, detailed GPS information and a list of every mobile phone number owned was collected for each household in the sample. The field team used a variety of different methods to reduce attrition as much as possible. First, up to three revisits were made to each household. Field supervisors were in regular phone contact with community leaders in order to help organise feasible interview times for revisits. Second, the survey team conducted extensive 'sweeping' efforts to revisit households on multiple occasions in the event that a household or household member was not available. As a part of the sweeping efforts, teams of enumerators attempted to revisit households that could not be reached initially while they were still in the district conducting interviews. Mobile phone numbers were collected when possible during the first visit, which sometimes enabled the enumerators to set up interview times directly with the households. A separate 'anthropometric sweep' was carried out toward the end of fieldwork to re-collect as many of the measurements missed during this initial sweep as possible. After all villages had been visited, the last week of fieldwork was devoted to a large team-wide sweeping exercise to interview initially unavailable households or household members. During this sweep the field team successfully interviewed an additional 169 households from across 146 villages (including 33 non-sample villages that some of our households had moved to). Third, OPM retained a dedicated sweep person, whose job was to conduct revisits. The ‘sweeper’ worked independently, using phone numbers and tracking information collected by enumerators during the first visit to visit as many households as possible. Last, phone interviews were conducted in the event that a male respondent could not be located in person. In every village, supervisors submitted a report to indicate the interview status for all sampled households.

In the end, attrition was higher than initially anticipated. Household residences changed more frequently than expected, with relocation to urban areas being especially likely. This was particularly true in the villages located closer to urban areas, where household mobility in search of economic activities was common. In some study areas, households followed customs that dictated that they send young children to live outside of the household for an extended period, often to live with grandparents.

Endline interviews were completed in 2,477 households, with 1,243 of those being in treatment villages and 1,234 in control villages. In addition to these, another 118 surveys were partially completed, which included households where the full interview with the primary female was conducted and anthropometry was measured, but no primary male was interviewed. Of the 238 households for which the survey was not even partially completed, 122 households had permanently moved to a location outside of Iringa region and could not be reached by phone, 102 households were long-term unavailable (typically due to short- or medium-term moves away from Iringa region), 12 households were not found, and two households did not give consent to be interviewed.

Table 4.3 presents the endline survey completion for the primary females and primary males in sample households. Substantially more men were interviewed at endline than at baseline, with 1,688 in-person interviews and another 516 conducted over the phone. Availability of primary females was lower, with the primary female in just 87.3% of baseline households completing an endline survey.

**Table 4.3: Primary females and males surveyed at endline, by baseline household status**

| | Primary female | | | | Primary male |
|---|---|---|---|---|---|
| | Pregnant | Child <12 months | Total | | |
| Surveyed at baseline | 948 | 1,885 | 2,833 | 100% | 1,418 |
| Available at endline | 845 | 1,627 | 2,472 | 87.3% | 1,688 |
| Phone interview | N/A | N/A | N/A | N/A | 516 |
| Caregiver interviewed at endline | 24 | 91 | 115 | 4.1% | N/A |
| Respondent unavailable/refused | 33 | 65 | 98 | 3.5% | 94 |

Source: Authors’ own

**Table 4.4: Attrition across encouraged and comparison arms**

| | Control mean | Treatment mean | Treatment effect | N |
|---|---|---|---|---|
| Baseline household interviewed at endline | 0.926 | 0.906 | -0.020 | 2833 |
| | | | (0.011)* | |
| Baseline primary female interviewed at endline | 0.955 | 0.949 | -0.006 | 2595 |
| | | | (0.009) | |
| Caregiver interviewed in place of primary female | 0.042 | 0.047 | 0.006 | 2595 |
| | | | (0.009) | |
| Primary male interviewed at endline | 0.846 | 0.853 | 0.008 | 2595 |
| | | | (0.015) | |

Notes: Estimates from the mNutrition Tanzania endline survey sample. Standard deviations are in parentheses and clustered at the village level. Treatment effect reports the coefficient on the treatment from an OLS regression of the outcome of interest on the treatment variable. The primary female is considered to be interviewed if the female from baseline was interviewed, while the primary male is considered to be interviewed if any male from a household was interviewed as the primary male at endline. * p<0.10 **p<0.05 ***p<0.01.

Source: Authors' own

Table 4.4 Presents the control and treatment group attrition rates as well as regression-based estimates of the relationship between the randomly assigned village-level mNutrition treatment status and the likelihood of different types of attrition. We find a small and marginally statistically significant negative relationship between the likelihood of household attrition and the mNutrition treatment. The point estimate indicates that households in treatment villages were two percentage points less likely to have been successfully interviewed at endline. Although the magnitude of this difference is small—just 2.2% of the control group completion rate—it is important to mention and address as we move forward in presenting and discussing the main impact estimates. In Annex C, we show the robustness of the main results to correcting for attrition by using the inverse of the predicted probability of having participated in the endline survey to re-weight the main specifications.

Conditional on having some endline information, there are no differences between households from the two treatment groups in the likelihood of having a completed primary female interview, a completed primary male interview, or having a caregiver interviewed in place of the primary female.

## 4.4 Data quality and cleaning

### 4.4.1 Data quality assurance

Data from the endline household surveys were collected using the Survey Solutions program on tablets. OPMT and IFPRI were careful to ensure the quality of the data collection. The general approach taken to assure the survey quality consisted of five main elements.

First, fieldwork supervisors were responsible for conducting interview observations, ensuring that all protocols were followed by their teams in line with the training. They also reviewed interview files before they were submitted to the server to check for issues and discussed any issues they identified with enumerators. Second, the project management staff, the data manager, and the field manager accompanied the teams in the field during the initial stages of the survey roll-out. They observed interviews, checked data, and provided feedback to give extra support in this early period of fieldwork. Third, OPMT programmed a series of checks within the CAPI questionnaire itself to flag possible errors to enumerators during the interview. This included checks on answers falling outside the expected range or logical inconsistencies. Fourth, the data manager ran a range of checks and monitoring procedures on a constant basis during fieldwork to rapidly identify any issues with the data and respond appropriately. The checks were divided between a series of Stata-based procedures to monitor any inconsistencies or indicators that would not be directly checked within CAPI validation conditions, as well as a bespoke dashboard, built by the data manager using the Power BI software, to monitor sample implementation, the progress of the survey, and individual interviewer performance by reference to a series of indicators and data visualisations tailored specifically for this survey.

Finally, additional visits to all field teams were conducted later in the fieldwork by the fieldwork manager, deputy project manager, and data manager to carry out additional observation of live interviews, conduct back-checks, seek feedback from village and community leaders, and monitor the general progress of the fieldwork. The purpose of these visits was to resolve issues that had been identified in the data and to verify the reported unavailability of some households. The survey management team conducted field monitoring visits, focusing on villages where interviews had seemed to have higher than average levels of inconsistencies. Altogether, the survey management team carried out quality assurance visits to 19 villages in Kilolo, Mufindi, and Iringa districts.

### 4.4.2 Data management

At the end of each day, interview files were checked by the enumerator, marked as complete, and then synchronised to the Survey Solutions server from the enumerator's tablet. The data manager downloaded the data regularly from the server. Raw data were stored on the server and the raw downloaded data were stored in a shared Dropbox folder. The data manager reviewed the data each day and ran a Stata checking do-file over the data to identify errors or inconsistencies. He communicated regularly with enumerators and supervisors to discuss any challenges they were facing and provided feedback on the submitted data in case of any errors. He also liaised directly with respondents by phone, if needed, to address any inconsistencies, and advised enumerators to conduct revisits to address errors or inconsistencies. Any recurring errors or inconsistencies were discussed with supervisors during fieldwork, so that they could be discussed as a survey team.

If the data manager identified issues to be addressed by the enumerator, the individual interview file was returned to the enumerator's tablet. The Survey Solutions software allows the data manager to monitor updates to data files after submission and responses to comments by enumerators. Once the data had been corrected or confirmed by the enumerator, they were again sent to the server for review and approval.

Data checks aimed to identify any questions that were not answered but should have been, to identify any children missing from the relevant rosters, and to flag any inconsistent response patterns. Where data were found to be incorrectly missing, follow-up phone calls and in-person revisits to respondents were conducted to collect the missing data. The data manager regularly checked for the completeness of the data from each village, running checks to ensure that all sampled households had been interviewed correctly and that all sampled households had been tracked successfully. For households that had not been tracked, the data manager parsed the provided reasons for why the household was not interviewed and checked when revisits were scheduled. As a part of this process, the data manager also checked for outliers in the data and confirmed these values by phone with either the enumerator or the respondent.

# 5 Experience with the mNutrition service

In the following section we discuss the findings from the modules of the endline survey intended to measure subjects' exposure to the mNutrition service and their self-reported satisfaction with the service and the content contained therein. The module was administered to the primary female and the primary male in all households, regardless of their treatment arm. Data collection on interaction with the service began by eliciting information from the primary female and the primary male (if available) about receipt of any text messages containing nutrition information during the previous two years. If individuals reported receiving nutrition-related text messages, they were then asked if they knew who sent the content. To decrease the likelihood of under-reporting due to recall issues, respondents who initially reported not having received any such messages during the previous two years were subsequently asked directly whether they had received the mNutrition messages during this period.

When available, we use the combination of the responses from the primary male and the primary female to generate a household-level measure of service exposure. Specifically, our preferred measure of service exposure is equal to one if either the primary male or the primary female reported receiving, or reported that their spouse received, automated text messages with nutrition information during the previous two years, and zero otherwise. We also use the information elicited on the sender of the messages to generate analogously constructed indicators of whether the household received nutrition messages from different senders during the previous two years. The potential senders included the organisation/number from which the mNutrition service content was sent; an NGO or international organisation; a government hospital, dispensary, or health centre; or a number associated with another service operated by the organisation that was sending out the mNutrition content. There was an option to report not knowing who the sender of the content was, and an option to report that the content was coming from some other source. We present exposure levels using all possible responses, but we select the broadest possible measure as our preferred exposure indicator as treatment households may have been likely to misattribute the content they received to other sources.

For individuals who reported having received mNutrition messages we followed up with a series of questions about their experiences with the programme—for example, how often the messages were received, whether they read the messages, and their satisfaction with the messages. While these measures are missing for any individuals who did not specifically report having received the mNutrition messages, it was not possible to elicit satisfaction or experiences with the mNutrition service from individuals who said they never received these messages.

## 5.1 Exposure and access to the service

Table 5.1, Table 10.1, and Table 10.2 present information on exposure by treatment status for households—aggregating across males and females using the combination of the individual responses—for the primary females and for the primary males, respectively. The difference in exposure between the treatment and control groups is immediately obvious: 66.4% of treatment households reported having received nutrition-related automated text messages during the preceding two years, as compared to 26.7% of control households. A t-test of the null hypothesis that there is no difference in the likelihood that treatment and control households received any nutrition-related text messages is firmly rejected (p-value < 0.001) and the normalised difference similarly suggests that treatment households were much more likely to have been exposed: the normalised difference of 0.868 is more than three times the 0.25 magnitude suggested by Imbens and Rubin (2015) as a rule of thumb for identifying imbalance.

Differences in reported exposure by respondent-identified sources also suggest that treatment households were substantially more likely to have been exposed to the mNutrition service. Treatment households were more likely to report having received nutrition messages from an NGO or international organisation (7.5% of treatment households, compared to 2.5% of control households; p-value < 0.001), more likely to report having received nutrition messages from the 15014 number that corresponds to another SMS-based programme run by the mHealth Tanzania-PPP (6.8%, compared to 2.0%; p-value < 0.001), more likely to report having received nutrition messages from an unknown source (26.3%, compared to 12.1%; p-value < 0.001), and more likely to report having received nutrition messages from the 15501 number that corresponds to the mNutrition service (42.7%, compared to 12.8%; p-value < 0.001). We highlight the differences in reported exposure for these other sources because it suggests that respondents may not have been able to correctly identify the source of nutrition text messages that they or their spouse received. For this reason, we base most of our subsequent discussion on the observed difference in exposure to any nutrition-related text messages in the previous two years.

With our preferred measure of mNutrition service exposure the implied gap between treatment and control households is 39.7%. This is well below the 70% take-up gap under which we conducted power calculations prior to the baseline survey (Barnett *et al*., 2017). If the self-reported measures of mNutrition exposure are accurate, we will have less statistical power to detect programme impacts. For example, for the HAZ outcomes, with a 70-percentage point take-up gap with the realised sample size of the study, the minimum detectable effect size would be a 0.25 standard deviation difference between the treatment and control groups. With the 39.7-percentage point gap in take-up suggested by the self-reported exposure measure the minimum detectable effect size increases to a 0.39 standard deviation difference in HAZ.

However, it is likely that the 39.7% gap in take-up is an underestimate of the true difference in exposure to the mNutrition service during the study period. Several months after the baseline survey was completed and the mobile phone numbers of treatment households had been registered for the mNutrition content we received information from the mHealth Tanzania-PPP on whether the full list of treatment group mobile phone numbers was registered for the mNutrition service content. The administrative mHealth Tanzania-PPP data suggested that between 85.9% and 98.3% of the treatment household mobile phone numbers were registered for the service in November of 2017, with the true rate dependent on what assumptions we make about the treatment phone numbers for which we did not receive a registration status report.[11] At 92.1%, the midpoint of the potential range in registration rates for the treatment group, there is little scope for upward bias in the self-reported service exposure measures. Therefore, if the reporting error is uncorrelated with true exposure, any measurement error is likely to generate a downward biased estimate of true exposure among treatment households, and similarly a downward biased estimate of the take-up gap. A symmetric argument for control households would generate an upward biased estimate of control group take-up and, again, a downward bias in the estimate of the take-up gap. We therefore think that it is likely that the 39.7% gap in self-reported exposure is likely to be a considerable underestimate of the true take-up gap. Using the lower bound for how many of the treatment numbers were registered for the mNutrition content as at November 2017 (85.9%) and the self-reported receipt of any automated nutrition messages for the control group (26.7%), or the self-reported receipt of nutrition messages from the mNutrition number (12.8%), yields an estimated take-up difference that is much closer to the 70% we assumed during power calculations: 59.2% in the former case and 73.1% in the latter.

[11] The ambiguity in registration rates for treatment households comes from the fact that we did not receive a registration status for 194 of the mobile numbers we requested information on. If we assume that these ‘missing’ numbers were not registered for the programme then the registration rate is 85.9%; if we assume that they were registered for the programme then the registration rate is 98.3%. The true registration rate is likely somewhere between these two bounds.

**Table 5.1: Household-level receipt of nutrition text messages by treatment status**

| | N | All | Treatment (T) | Control (C) | Normalised difference between T and C | P-value |
|---|---|---|---|---|---|---|
| Household received automatic text messages with nutrition information in past two years | 2,559 | 0.465 | 0.664 | 0.267 | 0.868 | 0.000 |
| | | (0.499) | (0.473) | (0.442) | | |
| Don't know if household receives nutrition messages | 2,559 | 0.001 | 0.000 | 0.002 | -0.056 | 0.154 |
| | | (0.028) | (0.000) | (0.039) | | |
| Household receives nutrition messages from government hospital | 2,559 | 0.010 | 0.010 | 0.009 | 0.008 | 0.835 |
| | | (0.098) | (0.100) | (0.096) | | |
| Household receives nutrition messages from NGO/international organisation | 2,559 | 0.050 | 0.075 | 0.025 | 0.232 | 0.000 |
| | | (0.218) | (0.264) | (0.156) | | |
| Household receives nutrition messages from church/mosque | 2,559 | 0.000 | 0.001 | 0.000 | 0.040 | 0.318 |
| | | (0.020) | (0.028) | (0.000) | | |
| Household receives nutrition messages from Wazazi Nipendeni/15501 | 2,559 | 0.277 | 0.427 | 0.128 | 0.708 | 0.000 |
| | | (0.448) | (0.495) | (0.334) | | |
| Household receives nutrition messages from 15014 number | 2,559 | 0.044 | 0.068 | 0.020 | 0.234 | 0.000 |
| | | (0.205) | (0.252) | (0.141) | | |
| Household receives nutrition messages from an unknown source | 2,559 | 0.192 | 0.263 | 0.121 | 0.367 | 0.000 |
| | | (0.394) | (0.440) | (0.326) | | |

Notes: Estimates from the mNutrition Tanzania endline survey sample. Standard deviations are in parentheses. P-values are from the tests of the difference in means between treatment and control group.

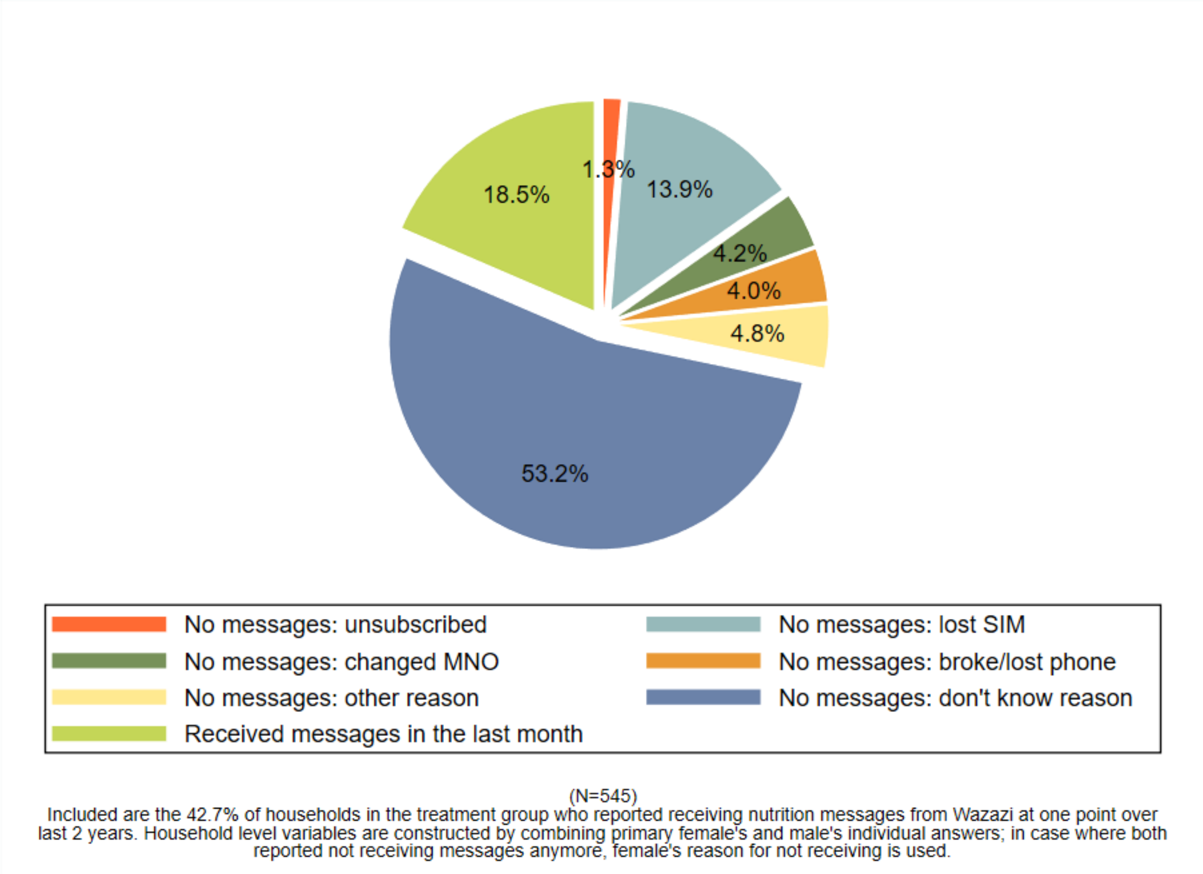
Source: Authors' own

Table 10.1 and Table 10.2 present the same exposure information separately for the surveyed primary females and primary males, respectively. The observed participation patterns are similar, regardless of which indicator is used. Thus, 48.5% of primary females in treatment villages reported having received some automatic nutrition text messages during the past two years while only 15.1% of primary females in control villages did so. The resulting gap in self-reported take-up among primary females is 33.4%. Among primary males the corresponding figures are 42.9% in treatment villages and 17.1% in control villages, yielding a take-up gap of 25.8%. These are similar, albeit slightly smaller, than the 39.7% gap in take-up based on the combination of male and female self-reporting.

While the statistics shown in Table 5.1, Table 10.1, and Table 10.2 capture the levels and differences across treatment groups in self-reported exposure to the mNutrition content at any point during the previous two years, it is also useful to characterise more recent receipt of the mNutrition content. To that end, Figure 5.1 plots the share of treatment households who reported ever receiving the mNutrition messages (from the 15501 number) that received a message in the month preceding the survey. For treatment households that identified as having received mNutrition messages in the previous two years but not in the last month, the figure disaggregates these households according to why they believe they had not received any content in the past month.

We see that 18.5% of the 545 treatment households that reported ever receiving the mNutrition content also received a message in the month preceding the survey. The remaining 81.5% did not receive any mNutrition content in the previous month. Of these, 56.3% did not know why they had not received a message, 15.2% had lost the SIM card that was registered for the service, 6.4% did not receive any message for some other reason, 4.2% had changed MNO, and for 4.0% their phone had either been broken or lost. Together, 23.4% of treatment households that had ever received messages had stopped receiving messages due to issues related to low mobile phone or mobile phone network attachment. Clearly, low attachment to mobile phones presents a potential barrier to effectively disseminating information through voice or text messages.

**Figure 5.1: Receipt of mNutrition messages in the last month, treatment group only**




Source: Authors' own

## 5.2 Interaction with the mNutrition platform

There are two methods for gauging individual interaction with the mNutrition service: the frequency with which messages were received on mobile phones and, conditional on having received a message, the likelihood that individuals read the content. Below we present self-reported measures of both types of service interaction for the primary female respondents and the primary male respondents, as well as through a household-level aggregation of the individual member responses. In cases where only one household member reported ever receiving mNutrition content, the household-level measures are identical to the individual-level measure for the ever-receiving household member. In cases where both the primary female and the primary male reported receiving mNutrition content during the previous two years, we construct the household-level maximum and minimum frequency of message receipt and the maximum and minimum likelihood of reading the message conditional on receipt. We also calculate the mean frequency of message receipt and likelihood of message reading as the midpoint of the interval between the value when using the household-level minimum and the value when using the household-level maximum.

The mNutrition message frequency was intentionally designed to vary depending on the pregnancy gestation and age of the child. In particular, messages should have been received less frequently as children aged. If respondents weight months just prior to the endline survey more than they weight earlier months—when sample children were younger or in utero—then the self-reported message frequency will underestimate the true frequency during the study period. Despite this, we still view the self-reports as providing meaningful information about interaction with the service.

### 5.2.1 Frequency of interaction

We first describe the regularity of messages for individuals in treatment households who reported ever receiving the mNutrition content or having a spouse who ever received the mNutrition content. Enumerators asked ever-receiving individuals ‘How often do you normally receive messages from the Wazazi Nipendeni/HPHB/15501 number?’ Respondents were allowed to respond ‘Every day’, ‘A few times a week’, ‘Once a week’, ‘A few times a month’, ‘Once a month’, ‘Less than once a month’, ‘Have only received one or two messages’, or ‘Don’t know’.

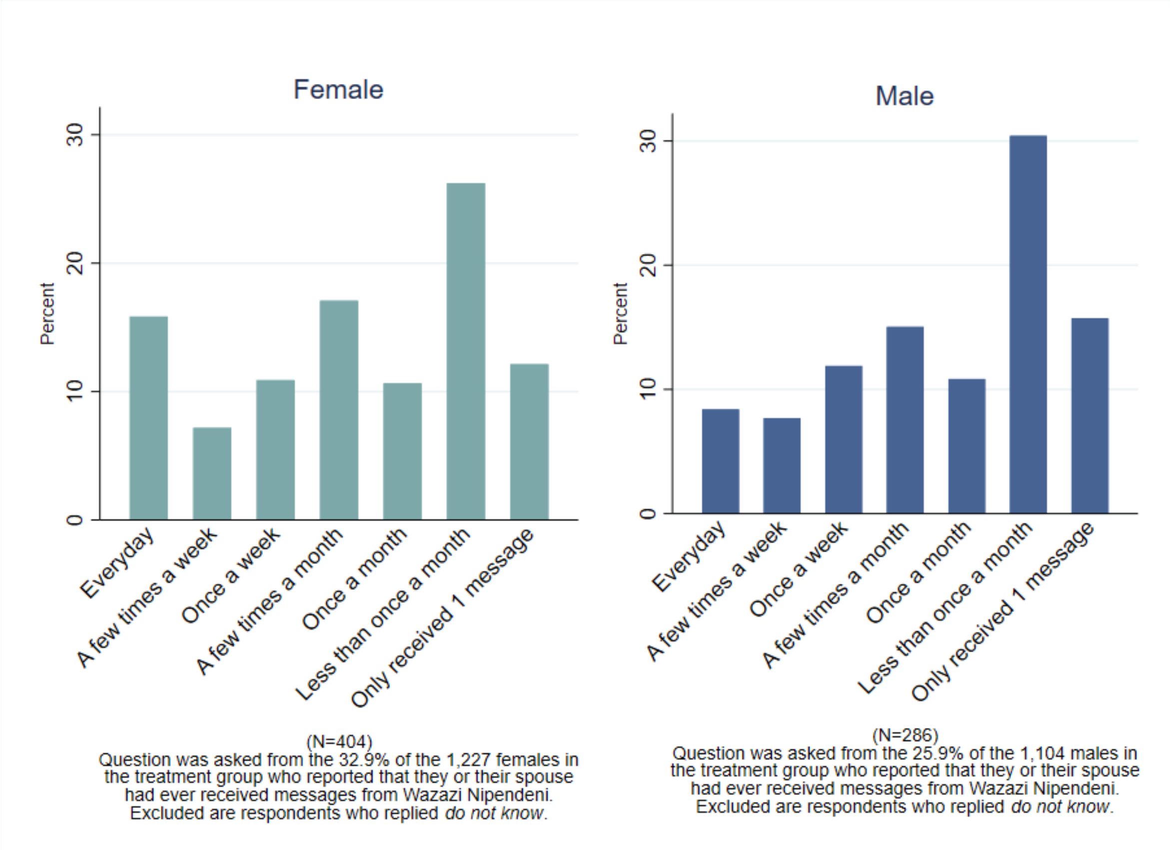
Figure 5.2 displays the frequency of mNutrition message receipt for the 32.9% of primary females and 25.9% of primary males from treatment households that reported ever receiving, or having a spouse who ever received, the mNutrition content based on the answers to this question. Over 35% of the ever-receiving primary females and primary males received messages less frequently than once per month. Among primary females, 26.2% received mNutrition messages less than once per month and 12.1% only received one or two messages; among primary males, 30.4% reported receiving messages less than once per month and another 15.7% only received one or two messages.

The remaining 61.7% of primary females and 53.9% of primary males received mNutrition content at least once per month. Among females, 15.8% received messages daily, 7.2% received messages a few times per week, 10.9% received messages once per week, 17.1% received messages a few times per month, and 10.6% received messages once per month. For males, 8.4% received messages daily, 7.7% received messages a few times per week, 11.9% received messages once per week, 15.0% received messages a few times per month, and 10.8% received messages once per month. The individual-level reports on message frequency therefore suggest a range of interaction levels with the mNutrition service among treatment households, with nearly half of ever-receiving individuals exposed to the content less than once per month.

Figure 5.3 plots the household-level aggregate of the primary female and primary male reports on message frequency. Unsurprisingly, the pattern is consistent with the individual-level results shown in Figure 5.2, regardless of whether the minimum or maximum statistic is used: a large share of ever-receiving households reported message receipt less frequently than once per month (36.6% using the maximum and 46.2% using the minimum) and the remaining households are distributed roughly evenly across the once a month, a few times a month, once a week, a few times a week, and daily frequencies. Importantly, these message receipt frequencies are for the 42.7% of treatment households that reported ever receiving the mNutrition content. Combining the information on the extensive margin of mNutrition receipt with the information on the frequency of message receipt indicates that, conservatively, between 23.0% and 27.1% of treatment households received mNutrition messages at least once per month during the study period. If, instead, we use the measure of whether the household ever received any automatic text message with nutrition information as a proxy for the extensive margin of service exposure[12] we would conclude that between 35.7% and 42.1% of treatment households received the service content at least once per month. Using 92.1%—the midpoint of the potential mNutrition registration rates from administrative data—as the measure of the extensive margin for service exposure would indicate that between 49.5% and 58.4% of treatment households received mNutrition content at least once per month. Thus, while even under conservative assumptions a meaningful share of treatment households received mNutrition messages at least monthly, this share roughly doubles if we use the administrative data from early in the study period.

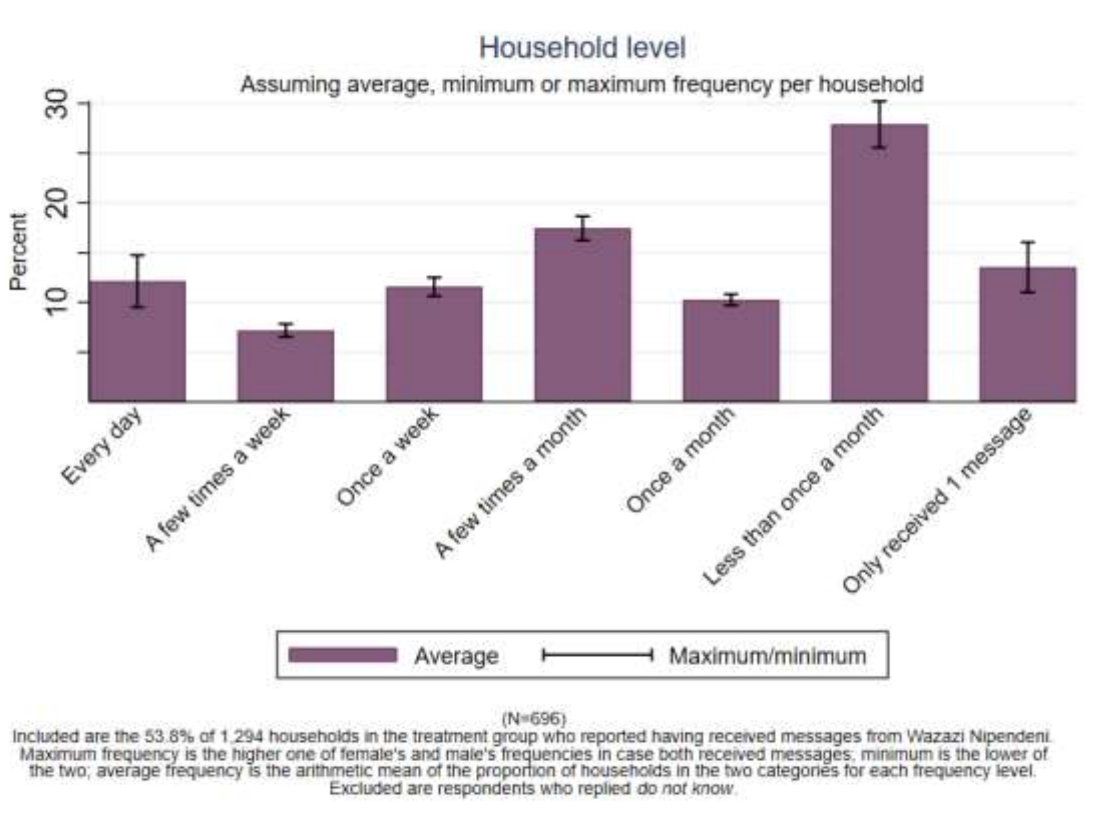
[12] We additionally need to assume that the distribution of the intensive margin of message frequency is the same for households that identified the mNutrition service as the source of the nutrition content and for households that did not.

**Figure 5.2: Frequency of receiving messages from mNutrition, individual level**




Source: Authors' own

**Figure 5.3: Frequency of receiving messages from mNutrition, household level**




Source: Authors' own

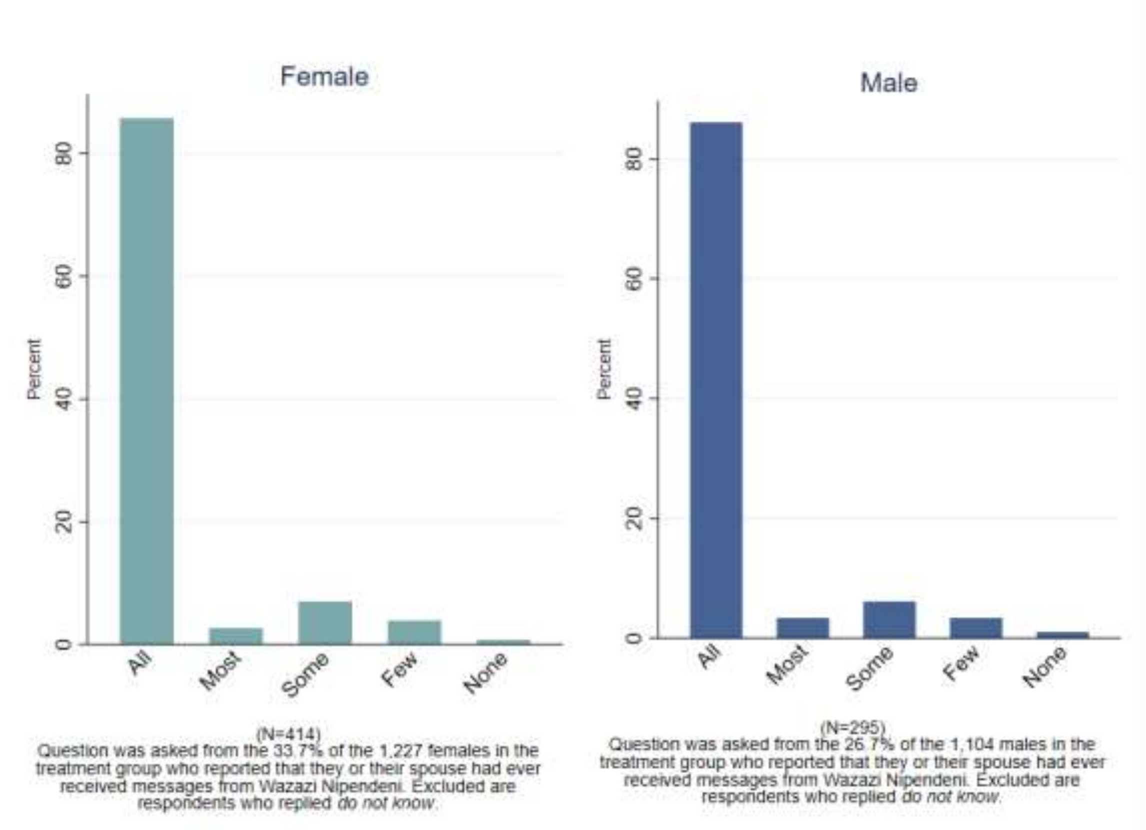
### 5.2.2 Frequency of reading mNutrition messages

A second potential barrier to successfully dispersing information through the mNutrition service is if individuals who receive the mNutrition content on their mobile phones do not read the messages. To better understand whether this is likely to be a binding constraint for the households in our sample, ever-receiving individuals were asked 'How many of the (mNutrition) messages you receive do, or did, you read?' The potential answers were 'All', 'Most', 'Some', 'Few', or 'None'. Figure 5.4 and Figure 5.5 display the reported frequencies based on the responses to this survey question. Figure 5.4 shows the frequency separately for the primary females and primary males from treatment households that ever received the mNutrition content and Figure 5.5 displays the household-level frequency using the aggregation method described above to generate the minimum, maximum, and mean statistics.

The results indicate that household members in treatment households were overwhelmingly likely to read the mNutrition content they received on their phone. Thus, 85.8% of ever-receiving primary females and 86.1% of ever-receiving primary males reported reading all the mNutrition messages they were sent. Another 2.7% of ever-receiving primary females read most received messages, 7.0% read some, 3.9% read a few of the messages, and 0.7% read none of the messages. Among the primary males, 3.4% read most of the messages, 6.1% read some, 3.4% read a few of the messages, and 1.0% read none of the messages.

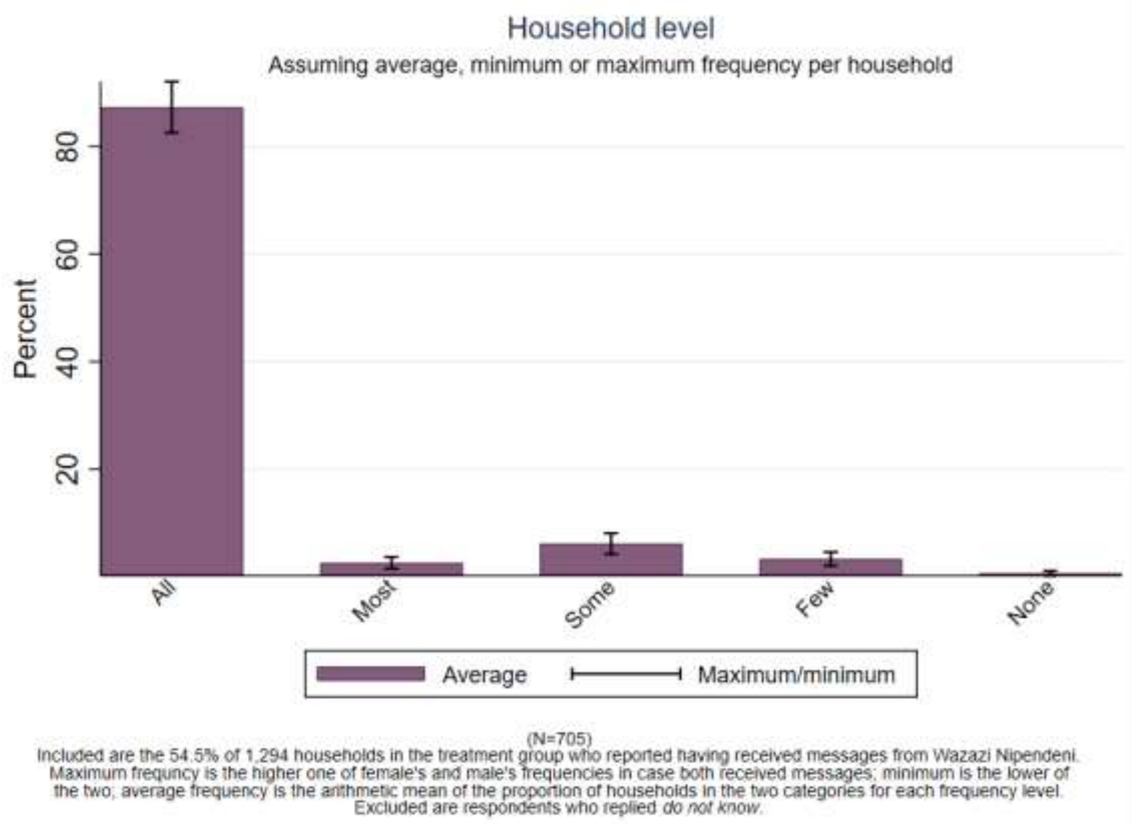
The household-level aggregate measures mirror the patterns in the individual-level data. Between 82.7% and 91.6% of households read all the mNutrition content they received. Not reading any of the received mNutrition messages was extremely uncommon, with between 0.1 and 1.3% of households falling into this category. Thus, while ensuring that households and individuals are receiving the mNutrition messages with regularity appears to be a relevant constraint, the households in our empirical sample are extremely likely to read the messages that are successfully sent to their mobile phones.

**Figure 5.4: Frequency of reading mNutrition messages, treatment group individuals**




Source: Authors' own

**Figure 5.5: Frequency of reading mNutrition messages, treatment group households**




Source: Authors' own

## 5.3 Perceptions of the mNutrition service

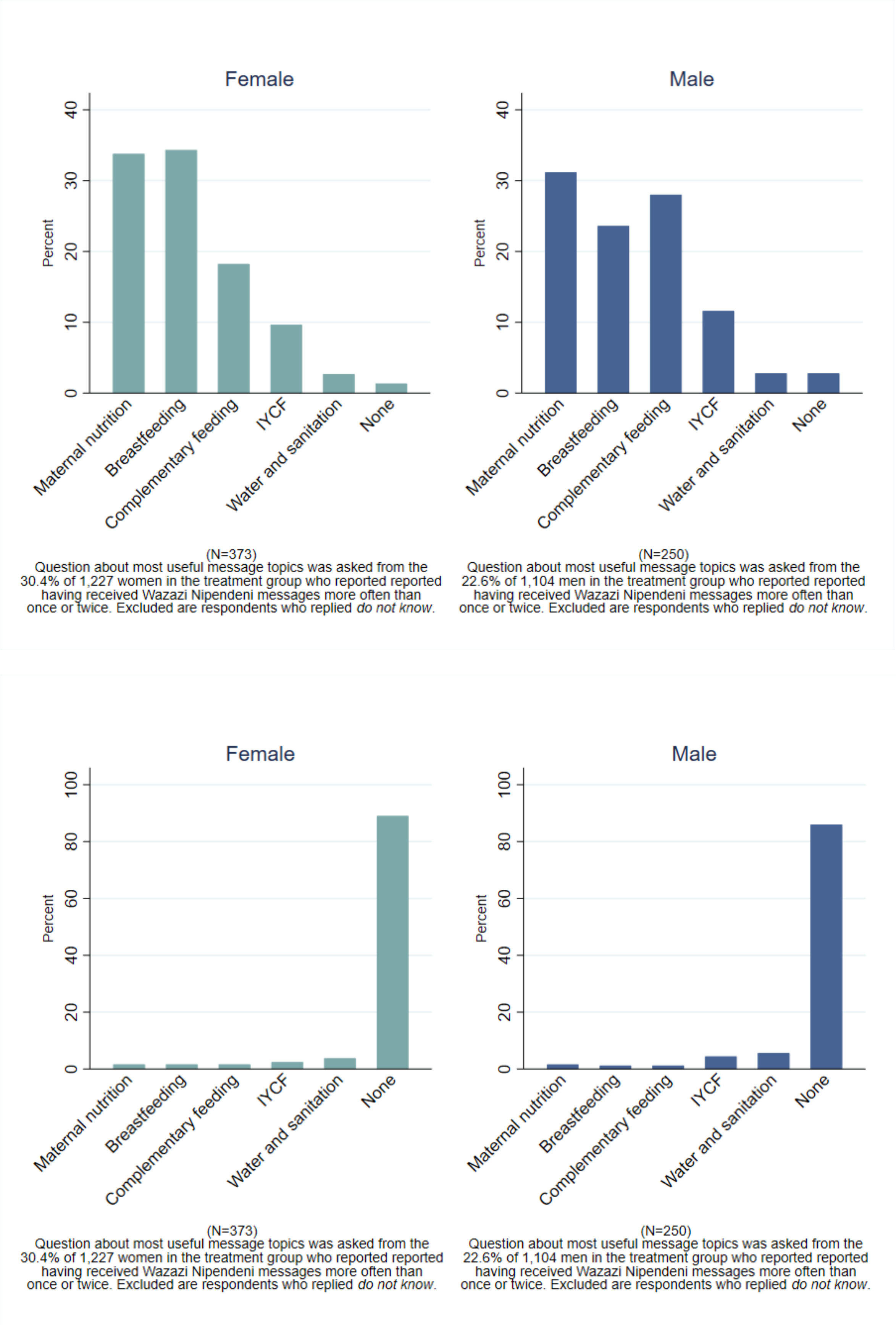
In addition to eliciting information on different dimensions of exposure to the mNutrition service, the endline survey also asked respondents who received the mNutrition content more often than once or twice during the previous two years to directly report on the perceived usefulness of the mNutrition service and the likelihood that they would recommend the service to others, as well as about which message topics they found to be the most and least useful. For the latter, respondents were asked 'What topics of messages do you find most useful?'[13] Enumerators read aloud the potential answers ('Pregnancy and maternal nutrition', 'Breastfeeding', 'Complementary feeding', 'Other IYCF messages', and 'Water and sanitation messages') and allowed respondents to identify the topic they found most useful. If respondents did not report finding any of the topics useful, then they were categorised as having 'Mentioned none'. Similarly, respondents were asked 'What topics of messages do you find least useful?' and read the same set of potential responses. Finally, we asked subjects whether they had put any of the tips they received through the mNutrition service into practice and, if they had, which category of recommendation (using the same potential response categories as the above questions) they had implemented.

Figure 5.6 shows the message topics that the primary female and primary male respondents found most useful and least useful. Interestingly, the self-reported ordering of message topic usefulness was similar across genders. Females found 'Breastfeeding' (34.3%) to be the most useful, followed by 'Maternal nutrition' (33.8%), 'Complementary feeding' (18.2%), 'Other IYCF' (9.7%), and 'Water and sanitation' (2.7%). Males found 'Maternal nutrition' to be the most useful topic (31.2%), followed by 'Complementary feeding' (28.0%), 'Breastfeeding' (23.6%), 'Other IYCF' (11.6%), and 'Water and sanitation' (2.8%).

Both females and males were reluctant to identify any message topic as not being useful, with 89.0% of females and 86.0% of males suggesting that none of the topics was the least useful. Roughly half of the remaining individuals indicated that the 'Water and sanitation' content was the least useful, with the residual responses distributed evenly across 'Maternal nutrition', 'Breastfeeding', 'Complementary feeding', and 'Other IYCF', but with these categories always containing fewer than 5% of the individuals who responded.

[13] The question about the general perceived usefulness of the mNutrition content, as well as about which topics were the most and least useful, were only asked of males or females who themselves reported having received the mNutrition content. The sample therefore varies slightly from the ever-receiving sample, which includes males and females with a spouse who received the content. We continue to describe these individuals as ever-receiving males and females for ease of understanding.

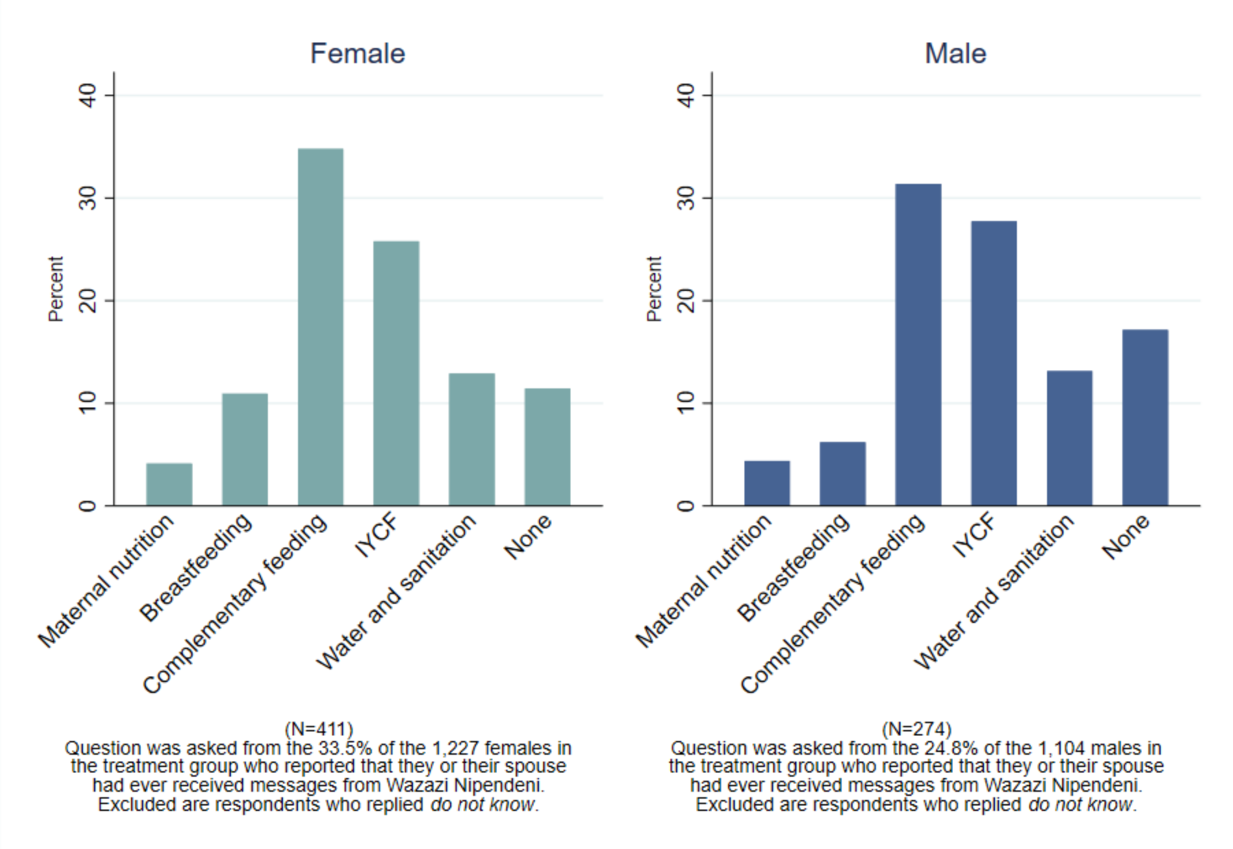
**Figure 5.6: Most and least useful mNutrition message topics, by gender**




Source: Authors' own

Consistent with the previous evidence that subjects typically found the mNutrition content to be useful, 88.7% of females and 81.0% of males reported implementing at least one tip they received through the service. Figure 5.7 plots the type of information that respondents reported implementing, with those who implemented none of the tips included as a separate category. For both females and males, complementary feeding and IYCF practices were the most likely messages reported to have affected behaviour. Interestingly, maternal nutrition and breastfeeding, which were highlighted as having been the most useful topics, were the least likely to have been put into practice. This potentially suggests that there are important barriers to adoption of these behaviours, which the dissemination of information alone is not able to overcome.

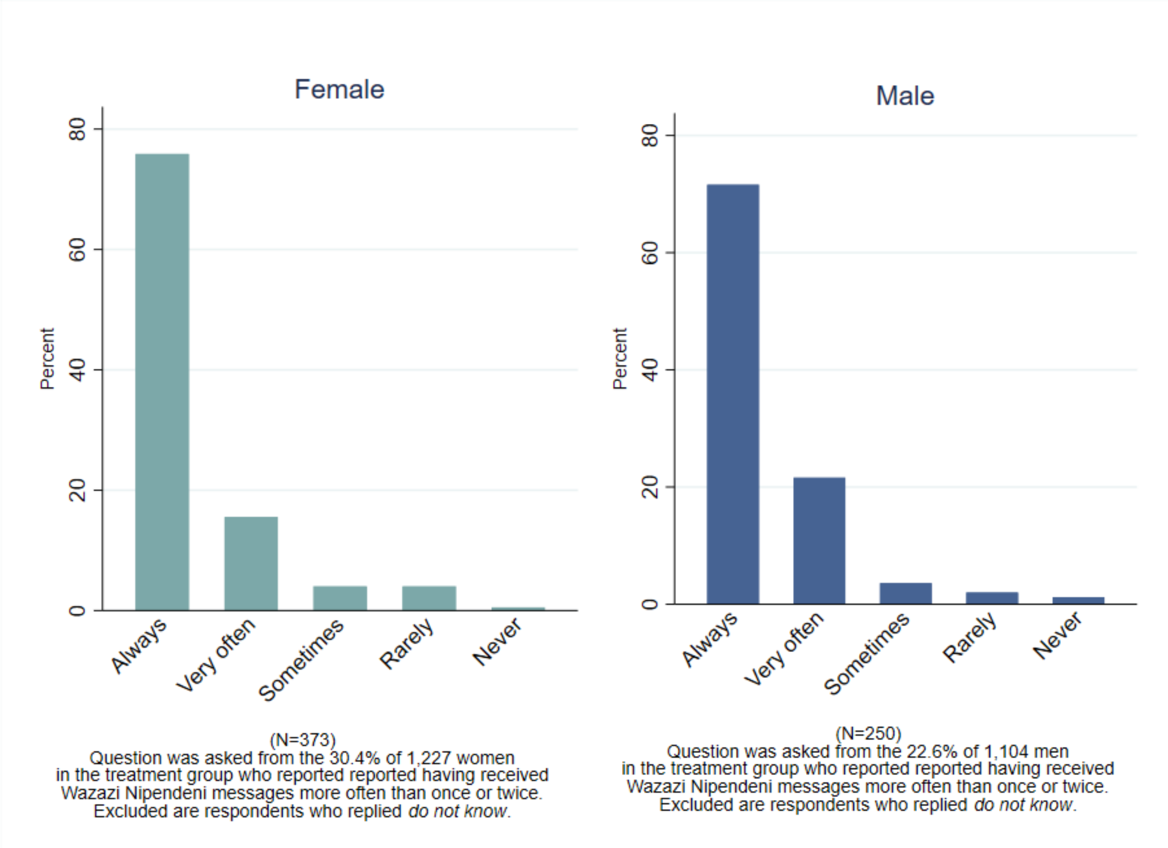
**Figure 5.7: Tips from mNutrition content that respondents have put into practice, by gender**




Source: Authors' own

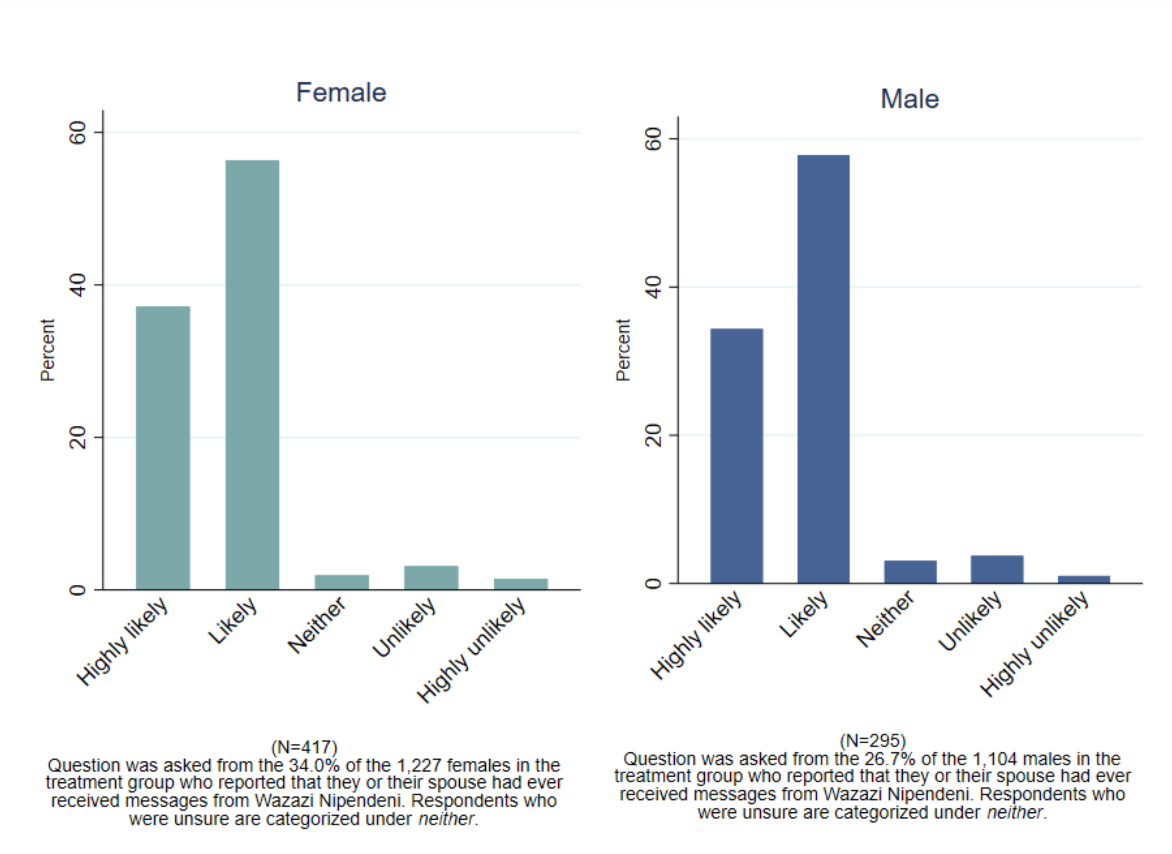
Figure 5.8 continues by displaying the frequency with which respondents found the mNutrition content to be useful. In this vein, 91.4% of females and 93.2% of males either always or very often found the messages useful. Just 4.6% and 3.2% of females and males reported rarely or never thinking the mNutrition messages were useful. Finally, Figure 5.9 plots the likelihood that respondents would recommend the mNutrition service to others. Here, 93.5% of females and 92.2% of males reported being likely or highly likely to recommend the service, while just 1.4% of females and 1.0% of males indicated they were highly unlikely to recommend the service. Together, both figures continue to suggest a high level of satisfaction with the content in the mNutrition messages among households that reported having been exposed to the service.

**Figure 5.8: How often was the mNutrition content useful?**




Source: Authors' own

**Figure 5.9: Likelihood of recommending the mNutrition service to others**




Source: Authors' own

## 5.4 Summary

Relative to households in control villages, households in villages randomly assigned to the mNutrition evaluation treatment group were 39.7 percentage points more likely to report having received any automatic text messages with nutrition information in the past two years (66.4% compared to 26.7%) and 29.9 percentage points more likely to report having received the mNutrition messages in the past two years (42.7% compared to 12.8%). Although statistically significant, these gaps are small relative to the take-up gaps we anticipated when conducting power calculations. If the smaller-than-expected difference in self-reported exposure to the mNutrition service is accurate, it will result in lower statistical power to detect significant differences in the outcomes between the treatment and control groups. However, administrative data on the registration rate of mobile phone numbers for treatment households early in the study period indicate that the true gap in exposure between the two groups may be considerably larger than the gap based on self-reports, and instead approach the 70 percentage point difference we anticipated when designing the evaluation.

Among individuals from treatment households who reported ever receiving the mNutrition content, self-reported message frequency varied widely across respondents. The key finding here is that 38.3% of primary females and 46.1% of primary males reported receiving mNutrition messages less than once per month. Conversely, 33.9% of primary females and 28.0% of primary males received mNutrition messages at least once per week during the study period. To the extent that these self-reported frequency of exposure measures are accurate, the differences in message receipt frequency will have important implications for the potential for households to absorb the information contained in the mNutrition content.

In contrast to the heterogeneity in reported mNutrition message receipt frequency, individuals who reported ever receiving the content were much more likely to agree about how often they read the messages they received. Thus, 85.8% of primary females and 86.1% of primary males reported reading all the mNutrition messages that they received on their mobile phones, with another 9.7% of females and 9.5% of males indicating that they read at least some of the content they were sent. Thus, at least for the households in the study sample, individuals were likely to read the mNutrition content that they were aware of receiving during the study period.

Based on the self-reported message receipt and reading frequencies discussed above, the consistent receipt and salience of received mNutrition content is the most probable barrier to successful dissemination of health and nutrition information through the mNutrition service. Exploring ways to increase message salience or to ensure the regular delivery of content, along with assessing the implications for the effectiveness of mobile phone-based interventions, are potentially fruitful avenues for future research.

Perceptions of the content in the mNutrition messages were nearly universally positive: 91.4% of females and 93.2% of males found the messages either always useful or very often useful. Similarly, 93.5% of females and 92.2% of males reported being likely or highly likely to recommend the service. Together with the high frequency with which individuals reported reading received mNutrition messages, the content appears to be both useful and appropriately adapted for the rural households included in the sample.

When asked to identify which message topics they found the most useful and which topics they were most likely to implement, males and females largely agreed. Respondents of both genders indicated that content on maternal nutrition, breastfeeding, and complementary feeding were the most useful, though they reported only being likely to implement tips related to complementary feeding. Breastfeeding and maternal nutrition-related suggestions, despite their perceived usefulness, were the least likely topics to have been implemented. This could, in part, be driven by important non-information-related barriers constraining behaviour for the women and households in the sample. For example, women in the sample who are engaged in wage labour may be unable to continue exclusive breastfeeding of children aged 0–5 months if they lack access to equipment to pump and properly store breast milk, or if they are not permitted to bring infants to work and take breaks for feeding. Receipt of information about the recommended length of exclusive breastfeeding would not address the binding constraints to exclusive breastfeeding for these women.

# 6 Main mNutrition impact estimates

In this section we present and discuss the main impact estimates for the impact of the randomly assigned offer of access to the mNutrition messages on the primary and secondary outcomes. As outlined in Section 3, we focus on the ITT estimates that rely only on the random assignment of mNutrition access across villages for identification. For outcomes where a baseline measure is available (child anthropometry, knowledge, women's dietary diversity, and household dietary diversity) we base our inference on ANCOVA models, which allow us to reduce the residual variance and lead to more precise treatment effect estimates by controlling for the baseline outcome measure.[14] For the highly age-dependent IYCF practice outcomes, which are only defined for children in a specific age range, we rarely observe the outcomes at both baseline and endline.[15] Impact estimates for these outcomes therefore rely on a simple differences estimator.

For the child-level outcomes related to anthropometry and IYCF practices we collected information on the focus child as identified in the baseline survey, as well as on any younger children born to the primary female. When available, we include information from these younger siblings in our analysis as they had similar potential to be exposed to knowledge and behaviour changes that resulted from the mNutrition intervention.

The ITT treatment effects estimate the impact of the random offer of access to the mNutrition service on the outcomes, regardless of whether households received the mNutrition content during the two-year study period. In our context, where baseline access to the intervention was low and nearly all treatment respondents offered access to mNutrition accepted the offer, we feel the ITT estimand is the most relevant for assessing the success of the service. In Section 7 we turn to alternative estimands—i.e. the LATE for compliers—that account for the observable differences in mNutrition take-up.

For all outcomes, we present the endline mean for the comparison group alongside ANCOVA or simple differences-based impact estimates and standard errors for the mNutrition service and the number of observations included in the analysis for that outcome. We show two estimates for each outcome. The 'basic controls' estimates include a control for whether the household had a pregnant woman at baseline (as opposed to having the mother of a child under 12 months at baseline). For the anthropometry outcomes the basic controls also include an indicator for whether the child was male, the child's age in months, and an indicator for whether the child being measured was the focus child (as opposed to the younger sibling of the focus child). For the IYCF practices outcomes, the basic controls also include the indicator for whether the child was male and the child's age in months. For household-level IYCF knowledge measures, we add indicators for whether the male had a knowledge score at endline and whether the female had a knowledge score at endline, as a number of households only had one respondent for the IYCF knowledge section.

A second set of estimates include a set of 'extended controls.' For child- and household-level outcomes these are household size at baseline, indicators for the district of residence at baseline, an indicator for whether the household was female-headed at baseline, an indicator for whether the household head was literate in Swahili at baseline, the household's predicted probability of

[14] In cases where an outcome is available at baseline and endline but the baseline value is missing for some individuals, we still employ ANCOVA. To do so, we code the baseline value for these missing observations to be zero and include an indicator for whether each observation had a missing baseline value. In this way we use all of the available baseline information to reduce the residual variance of the outcome without discarding observations or affecting the estimated relationship between the baseline and endline values of the outcome.
[15] For example, a child who was aged 6–8 months at baseline and therefore had a non-missing value for the indicator measuring whether children aged 6–8 months are receiving solid, semi-solid, or soft foods at baseline will be aged 30–32 months and have a missing value for that indicator at endline.

being below 150% of the national poverty line in Tanzania based on the Poverty Probability Index (PPI) score at baseline,[16] and whether the primary female owned a mobile phone at baseline. For the female-level outcomes the extended controls are the same except that the indicator for whether the household head was literate in Swahili at baseline is replaced by an indicator for whether the primary female was literate in Swahili at baseline, and a control for the primary female's age in years is also added. For the male-level outcomes the indicator for primary female ownership of a mobile phone at baseline is replaced by the analogous measure for the primary male, and a control for the age of the head of household is added.

In the estimates presented in this section standard errors are clustered at the village level (CRSE) to allow for any arbitrary within-village correlation in the errors. We display one, two, and three significance stars for p-values below 0.1, 0.05, and 0.01, respectively. We additionally show the CRSE-based p-values and randomisation inference-based p-values for key outcomes in Table 10.7 and Table 10.8 of Annex C. In Annex C we also show the robustness of the main results to correcting for attrition by using the inverse of the predicted probability of having participated in the endline survey to re-weight the main specifications.

The primary outcomes for the study are HAZ for children under the age of three, IYCF practices, and women's dietary diversity. The secondary outcomes are IYCF-related knowledge for females and males, as well as an exploration of whether sending the mNutrition content to the mobile phone of the primary male in addition to the primary female had a differential impact on the other primary and secondary outcomes. We present results and discuss the evaluation methods in more detail for this last secondary outcome in Section 8.1.

## 6.1 Primary outcomes

### 6.1.1 Child anthropometry

To preview the main results, we begin by plotting the full distribution of endline HAZ and WHZ for all measured children by their mNutrition treatment group. Figure 6.1 shows the distribution of HAZ scores while Figure 6.2 does the same for WHZ scores. In both figures we also plot a vertical line at the cut-off for stunting (in the HAZ case) or wasting (in the WHZ case).

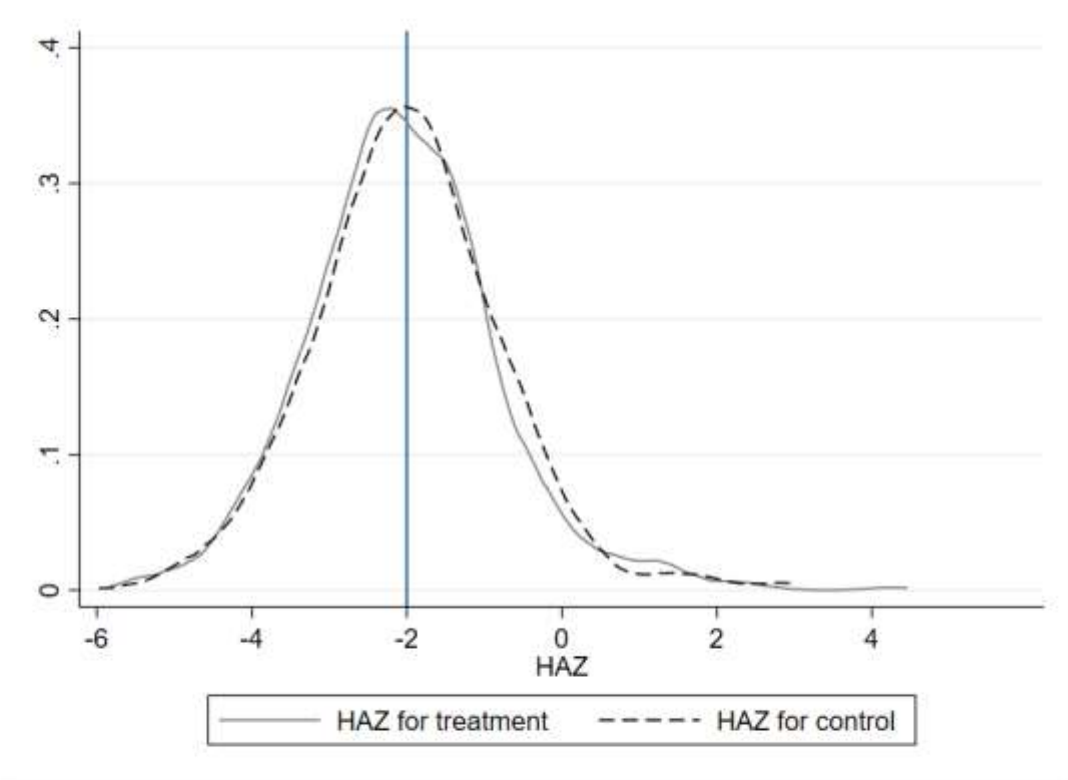
As was the case at baseline, the HAZ distributions for children in treatment villages and children in control villages track one another closely. The peak of the density for treatment children occurs slightly before the peak among control children, which could indicate a slight height disadvantage for treatment children around the stunting cut-off (HAZ < -2); however, the distribution for treatment children crosses below the control distribution around -1, suggesting—if anything—a slight advantage for treatment children beginning at a HAZ score of -1. Thus, at most there are small differences in the HAZ distribution across treatment groups, and the differences do not consistently favour one group.

Similarly, the WHZ distributions for treatment and control children are difficult to distinguish from one another, with at most small and inconsistently signed differences appearing. Control children are slightly advantaged relative to treatment children around the wasting threshold, but this ordering is reversed between -2. Few children are overweight (WHZ > 2), and there are no apparent differences in the density of treatment and control children above this point.

[16] PPI scores map households in relation to the likelihood that they will fall below different national and international poverty lines. We use the likelihood that households will fall below 150% of the national poverty line in Tanzania as our primary PPI score.

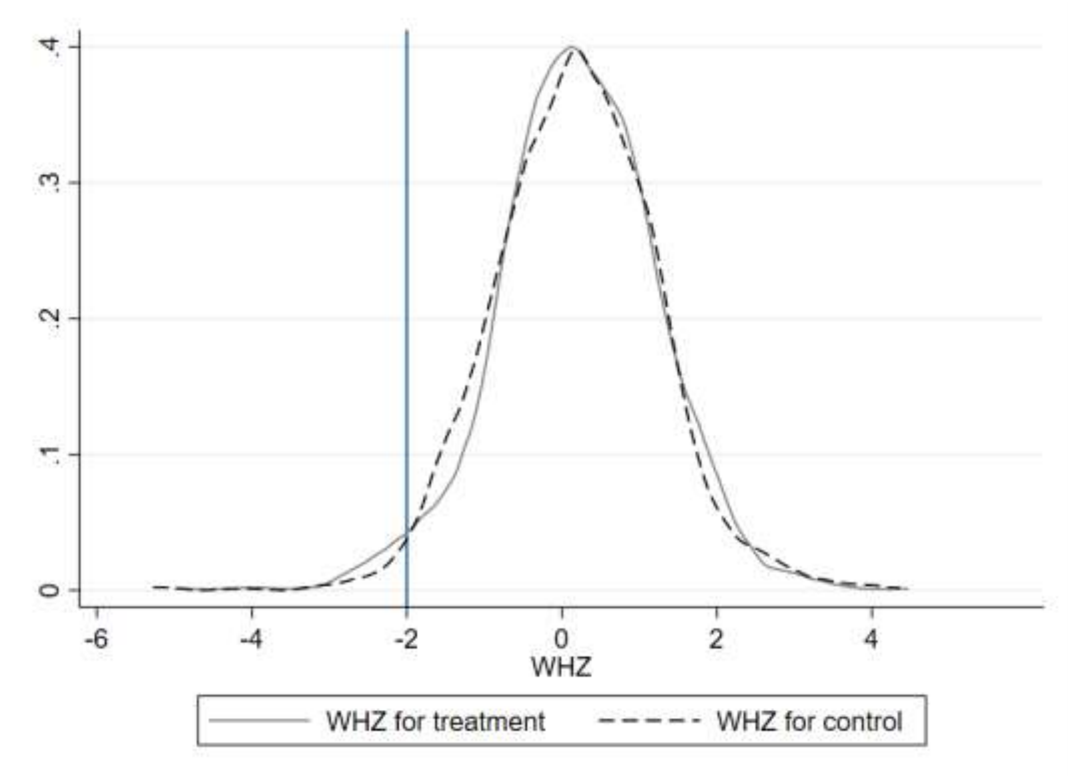
The raw HAZ and WHZ densities therefore suggest few, and at most minor, differences between treatment and control children. To explore potential impacts on anthropometry in a more parametric but systematic fashion, we next turn to the regression-based results described above.

**Figure 6.1: Distribution of children's HAZ by mNutrition treatment group**




Source: Authors' own

**Figure 6.2: Distribution of children's WHZ by mNutrition treatment group**



Source: Authors' own

Table 6.1 presents the main impact estimates for the child anthropometry outcomes. Consistent with the lack of observable differences in the HAZ and WHZ distributions across treatment groups, we find no evidence that the mNutrition service had any effect on HAZ or WHZ. The point estimate in both the basic and extended models for HAZ is negatively signed but small in magnitude (-0.06) and not statistically distinguishable from zero at the 10% level. The point estimate for WHZ is positive but even smaller than the HAZ coefficient (0.029) and similarly not distinguishable from zero. WAZ is likewise not affected by the offer of access to the mNutrition content. Unsurprisingly, given that the random assignment of mNutrition ensured balance in age and sex across the treatment and control groups, height (in cm) and weight (in kg), unadjusted for the age and sex of the child using the reference distribution of children, are also not impacted by the mNutrition service.

Table 6.1 additionally shows impact estimates for the likelihood that children are stunted, wasted, underweight (WAZ < -2), and overweight or obese (WHZ > 2). For none of these four outcomes are we able to reject the null hypothesis that there was no impact of access to the mNutrition service. The point estimates for stunting and wasting are positive but small in magnitude. At the 5% level, we can rule out any impact outside the 95% confidence interval, which includes all impacts larger than a 6.9 percentage point increase or a 1.2 percentage point decrease in stunting. By the same logic, we can rule out any impact on wasting larger than a 2.1 percentage point increase or a 0.2 percentage point decrease. Only 3.6% of the control group children are classified as overweight or obese, and, as mentioned, this is unaffected by access to the mNutrition messaging.

Across all the child anthropometry outcomes, we find no evidence that providing households with access to the mNutrition service had an impact—positive or negative—on any of the outcomes. In Annex C we show the simple difference analogues of the main ANCOVA estimates. The point estimates are nearly identical and in no cases are any of the conclusions different when using the two methods.

**Table 6.1: Impact estimates of impact of mNutrition on child anthropometry, ANCOVA**

| | Control mean | Impact estimates, basic controls | Impact estimates, extended controls | N |
|---|---|---|---|---|
| Child height (cm) | 79.593 | -0.153 | -0.075 | 2803 |
| | | (0.210) | (0.187) | |
| Child weight (kg) | 10.549 | -0.009 | -0.009 | 2803 |
| | | (0.062) | (0.062) | |
| Child HAZ | -1.976 | -0.060 | -0.039 | 2794 |
| | | (0.055) | (0.051) | |
| Child WAZ | -0.933 | -0.018 | -0.022 | 2802 |
| | | (0.043) | (0.042) | |
| Child WHZ | 0.182 | 0.029 | 0.005 | 2792 |
| | | (0.052) | (0.044) | |
| Anderson Index: HAZ and WHZ combined | 0.006 | -0.010 | -0.013 | 2800 |
| | | (0.031) | (0.030) | |
| Child stunted | 0.501 | 0.029 | 0.023 | 2794 |
| | | (0.020) | (0.019) | |

| | | | | |
|---|---|---|---|---|
| Child underweight | 0.167 | -0.015 | -0.014 | 2802 |
| | | (0.015) | (0.015) | |
| Child wasted | 0.015 | 0.010 | 0.010 | 2792 |
| | | (0.006) | (0.006) | |
| Child overweight or obese | 0.036 | 0.004 | 0.002 | 2792 |
| | | (0.007) | (0.007) | |

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 the treatment from an OLS regression of the outcome of interest on the treatment variable, controlling for baseline household classification, whether child is the focus child, child's age, child's gender, and value of the respective outcome at baseline. The latter is replaced with 0 for children who were not measured at baseline; indicators for missing measurements are included as controls. 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, and PPI score. Control mean is comparison group's mean at endline. * p<0.10 **p<0.05 ***p<0.01

Source: Authors' own

### 6.1.2 IYCF practices

Table 6.2 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, which frequently has little overlap with the age at endline of the focus children in our sample, we caution against placing too much weight on the results or non-results for these outcomes. This lack of overlap in age occurs because we purposefully targeted households with children aged 0–12 months of age at baseline, or with a pregnant woman at baseline. As a result, few households have a child aged 12–15 months, 0–5 months, or 6–8 months of age at endline as it requires the focus woman to have had a child after the focus child (the child 0–12 months or in utero at baseline), and for that birth to have occurred within a narrow time period.

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 food 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 children aged 6–23 months and children aged 6–35 months (Arimond *et al.,* 2010; Moursi *et al.,* 2008). As discussed in Section 3, we also show impacts on IYCF practice Anderson indices for children aged 6–23 months and 6–35 months.

Though positively signed, we find no statistically significant impact on the likelihood of early initiation of breastfeeding (point estimate 0.010 and p-value 0.41 in the basic controls case). Further, at the 5% level we 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 aged 0–5 months, and the consumption of solid, semi-solid, or soft foods for children aged 6–8 months, are negative, but based on only 165 observations and 58 observations, respectively, and not statistically distinguishable from zero.

We do observe 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. The point estimate suggests that the offer of access to mNutrition increased the likelihood that children aged 12–15 months were still breastfed by 26.2 percentage points (p-value: 0.057), or roughly 34% of the control group mean. As mentioned before, we are hesitant about placing too much weight on this result as it is driven by there being two control group children aged 12–15 months who were not breastfed, relative to no treatment group children in that age group not being breastfed.

Importantly, there are more observations for the count of the number of food categories children aged 6–23 months consumed and children aged 6–35 months consumed in the 24 hours preceding the survey (843 and 2,558). We find evidence that household access to mNutrition had a positive impact on dietary diversity for both age groups, although only the effect for children aged 6–35 months is statistically significantly different from zero. Access to mNutrition increased the number of food categories consumed by 0.107 (p-value 0.073) in the basic model and 0.112 (p-value 0.046) in the 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 (BCC) programme that included in-person counselling in Ethiopia (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 implemented in Burundi (Leroy *et al*., 2015).

We find 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. Point estimates suggest that treatment children are 20 percentage points (p-value 0.107) and 38 percentage points (p-value 0.087) more likely to have consumed vitamin A-rich fruits and vegetables or other fruits and vegetables in the basic models. However, the statistical significance of the service impacts depends on whether we use the basic or the extended model, suggesting that these effects are not robust. We also find a positive, albeit never statistically significant, impact on dairy consumption, with the point estimate indicating access to the mNutrition service is associated with a 3.0 percentage point increase (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 MDD standard (World Health Organization, 2008). Children in treatment villages were 3.8 percentage points more likely to have consumed from four food groups (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 we use 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 is not statistically significantly associated with the likelihood that they met the minimum meal frequency requirements, although the point estimate is positive (coefficient 0.025; p-value 0.343).

Meeting the MAD, the intersection of minimum meal frequency and MDD for children aged 6–23 months, is relatively rare among the control children: just 22.4% of the control group satisfy the MAD requirements. Children in the mNutrition treatment group are 6.9 percentage points more likely to satisfy MAD based on the model with basic controls and 7.8 percentage points more likely to satisfy MAD based on the model with extended controls. 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 6.2 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), although both point estimates are positive regardless of which model is used.

**Table 6.2: Impact estimates of impact 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 aged 0–5 months who are fed exclusively with breast milk | 0.032 | -0.014 | -0.016 | 165 |
| | | (0.027) | (0.024) | |
| Children aged 12–15 months who are fed breast milk | 0.846 | 0.262* | 0.302* | 27 |
| | | (0.131) | (0.168) | |
| Infants aged 6–8 months who receive solid, semi-solid, or soft foods | 0.697 | -0.050 | 0.045 | 58 |
| | | (0.130) | (0.151) | |
| Number of food groups (of seven) children aged 6–23 months consume | 3.685 | 0.056 | 0.055 | 843 |
| | | (0.093) | (0.089) | |
| Number of food groups (of seven) children aged 6–35 months 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 aged 6–23 months who consume four or more food groups | 0.609 | 0.001 | -0.001 | 843 |
| | | (0.034) | (0.033) | |
| Children aged 6–35 months who consume four or more food groups | 0.604 | 0.038* | 0.038* | 2,558 |
| | | (0.023) | (0.021) | |
| Children aged 6–23 months who meet the minimum meal frequency | 0.790 | 0.025 | 0.029 | 846 |
| | | (0.026) | (0.025) | |
| Children aged 6–23 months who meet the MAD | 0.224 | 0.069** | 0.078** | 830 |
| | | (0.034) | (0.033) | |
| Anderson Index: IYCF practices (children aged 6–23 months) | -0.003 | 0.007 | 0.010 | 1,011 |
| | | (0.007) | (0.007) | |
| Anderson Index: IYCF practices (children aged 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 the treatment from an OLS regression of the outcome of interest on the treatment variable, controlling for baseline household classification, child's age, and 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, and PPI score. Control mean is comparison group's mean at endline. * p<0.10 **p<0.05 ***p<0.01

Source: Authors' own

### 6.1.3 Dietary diversity

We present treatment effect estimates for two measures of women's dietary diversity in Table 6.3: the WDDS, which ranges from 0 to 10 depending on the number of food categories women consumed from in the day preceding the survey, and the MDD-W score, which is an indicator equal to one if women consumed from at least five of the 10 categories. Though it is not a primary or secondary outcome, we also show results for the Household Dietary Diversity Score (HDDS), as it helps to provide context for the two variables measuring children's dietary diversity show in Table 6.2 and the two variables measuring women's dietary diversity shown in Table 6.3.

The point estimates for both measures of women's dietary diversity are positive, suggesting that access to mNutrition may have improved the quality of women's diets, though only the impact on MDD-W is statistically distinguishable from zero at the 10% level. Access to mNutrition increases the WDDS by 0.073 in the basic model (p-value 0.279) and 0.076 in the extended model (p-value 0.249), although neither impact is statistically distinguishable from zero at the 10% level. For MDD-W, mNutrition increases the likelihood that the primary female satisfies the MDD-W by 4.0 percentage points in the basic model (p-value 0.062) and 3.6 percentage points in the extended model (p-value 0.076); the effect sizes for MDD-W represent increases of 6.4% and 5.8% of the control group mean. That the impact on MDD-W is larger than the impact on the WDDS suggests that access to the mNutrition service may have been more useful for women with low levels of dietary diversity that were below, or in danger of dropping below, the MDD-W threshold.

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). Interestingly, we find no evidence that the 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) relative to the control group mean and the corresponding coefficient for women, and does not approach 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 be driven by changes in the way that households allocate food across household members, rather than by affecting household-level consumption patterns more broadly.

**Table 6.3: Impact estimates of impact 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 | 0.039* | 0.035* | 2,535 |
| | | (0.022) | (0.020) | |
| HDDS (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 the treatment from an OLS regression of the outcome of interest on the 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, and PPI score. Control mean is comparison group's mean at endline. * p<0.10 **p<0.05 ***p<0.01

Source: Authors' own

## 6.2 Secondary outcomes

Table 6.4 turns to exploring whether mNutrition had any measurable impact on the knowledge of primary females and primary males regarding nutrition and IYCF practices. As mentioned in Section 3, we also show estimates for a household-level measure of knowledge that combines the male and female scores using the Anderson Index method (Anderson, 2008). Table 10.3 and Table 10.4 in Annex C present the impact estimates for the individual questions for females and males.

For both females and males, IYCF knowledge scores increased considerably between the baseline and endline: females in the control group answered 13% more questions correctly at endline, while control group males answered 10% more questions correctly. This could reflect the fact that the baseline knowledge assessment questions were asked again at endline, or it could simply be driven by knowledge acquisition as parents age and gain experience with children. Regardless, men still score lower than women—10% lower than women at endline, compared to a 5% difference at baseline in the control group.

Turning to the estimated impact of access to mNutrition, Table 6.4 shows that mNutrition increased knowledge of IYCF and nutrition for males, with a positive and statistically insignificant increase in IYCF knowledge for females. The point estimate for females suggests that the mNutrition content increased the percentage of questions answered correctly by 0.79% in the basic model (p-value 0.225) and by 0.74% in the extended model (p-value 0.223). The impact on male nutrition knowledge is more than twice the size of the female impact and statistically significant at the 5% level. Access to the mNutrition content increased the percentage of knowledge questions that men answered correctly by 1.71% in the basic model (p-value 0.018) and by 1.65% in the extended model (p-value 0.021).

There are several reasons why the mNutrition service may have affected male knowledge more than female knowledge. First, females had higher knowledge levels at baseline. To the extent that the content in the mNutrition messaging focused partly on the information that females were more likely to already possess than males, we should expect to see male knowledge increase more in response to the intervention. Second, the women of reproductive age who participated in the study may have been more likely to receive IYCF and nutrition information through interactions with other people or parts of the health system in Tanzania. For example, nutrition information could be disseminated to women through antenatal or postnatal check-ups, which nearly all the women in our sample attended at least some of. Conversely, men may have been less likely to have these types of interactions and their information acquisition in the absence of mNutrition may therefore have been more limited. Third, mobile phone ownership was importantly higher among males than females at baseline. Of the individuals re-interviewed at endline, 91.4% of males owned a mobile phone at baseline, as compared to 42.2% of females. As a result, for many treatment households the mNutrition messages were sent to a mobile phone owned by the primary male. If this information improves knowledge for the message recipient but the newly acquired knowledge is not passed without friction onto the other household member, then male knowledge would increase more, on average.

The last row of Table 6.4 shows that access to the mNutrition service also had a statistically significant impact on the combined household-level measure of nutrition knowledge. The point estimate indicates that treatment group households have a 0.091 standard deviation higher knowledge score than control households at endline if we rely on the basic model, and a 0.09 standard deviation higher knowledge score if we instead use the extended model. Both estimates are statistically significant at the 5% level, with p-values of 0.032 and 0.023, respectively.

**Table 6.4: Impact estimates of impact of mNutrition on females' and males' IYCF knowledge, ANCOVA**

| | Control mean | Impact estimates, basic controls | Impact estimates, extended controls | N |
|---|---|---|---|---|
| Percent of answers correct (female) | 68.969 | 0.793 | 0.739 | 2,469 |
| | | (0.651) | (0.605) | |
| Percent of answers correct (male) | 58.959 | 1.707** | 1.649** | 2,204 |
| | | (0.713) | (0.706) | |
| Anderson Index: combined percent correct: females and males | -0.047 | 0.091** | 0.090** | 2,114 |
| | | (0.042) | (0.039) | |

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 the treatment from an OLS regression of the outcome of interest on the 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, and PPI score. Control mean is comparison group's mean at endline. * $p<0.10$ ** $p<0.05$ *** $p<0.01$

Source: Authors' own

## 6.3 Summary

We find 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 more positive. There is some evidence that access to the mNutrition service improved dietary diversity for children: both the

Child Dietary Diversity Score (CDDS) and the likelihood that children satisfy the MDD threshold were statistically significantly increased by the mNutrition service for children 6–35 months. We also find that mNutrition improved the likelihood that children aged 6–23 months achieved a MAD. Similarly, women's diets in treatment households were more diverse, as measured by MDD-W, suggesting that treatment women were more likely to have nutritionally adequate diets. We observe 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.[17]

As expected given the observed impacts on dietary diversity, mNutrition also improved nutrition knowledge in treatment households. Although we estimate 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, roughly equivalent to answering 1.1% more questions correctly.

Together, this evidence suggests that the mNutrition messaging modestly improved the diets of women and young children in the study areas. The messaging had impacts on the nutrition knowledge of male household members, and these changes in information, or potentially a combination of changes in information and a strengthening of beliefs in existing information, generated differences in how children and women consumed food – specifically by increasing the number of food groups these populations consumed. That these changes in dietary diversity did not translate into decreases in stunting or improvements in HAZ for children suggests that the main barriers to early childhood growth failure in rural Iringa are not information- or knowledge-related, or that the changes in diet-related behaviours were not, on their own, large enough to meaningfully affect child nutritional status.

[17] We do not implement any adjustment for multiple hypothesis testing for the IYCF practices outcomes. Given the number of tests for the aggregate IYCF practices indicators (10, including CDDS and MDD for children 6–23 months and 6–35 months), we should expect to see approximately one difference at the 10% level by chance.

# 7 mNutrition LATE impact estimates

In this section we extend 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). Section 3.3.2 set out why we prefer the ITT estimates presented in the previous section, based on the fact that they are policy-relevant, easy to interpret, and require no assumptions beyond the conditional independence assumption that randomisation ensures is satisfied. However, we 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 also be a policy-relevant reflection of the benefits of the service.

In what follows we present the LATE estimates using 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.[18] To do so, we 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 Section 3.3.2, 2SLS estimates that use the random mNutrition offer as the excluded instrument identify the causal impact of having received the mNutrition content for complier households.

## 7.1 First-stage estimates

To begin, Table 7.1 shows the observation counts, estimates, and a diagnostic of the strength of the first-stage relationship: that is, of the relationship between the random offer of the mNutrition service to treatment households and the self-reported measure of whether households received any automatic nutrition messages during the study period. The diagnostic—the Kleinbergen-Paap Wald rk F-statistic—is a measure of the strength of the relationship between the excluded instrument and the endogenous indicator of nutrition messaging receipt, generalising the standard F-statistic to a context with clustered standard errors and potentially weak instruments to assess whether the instrument is sufficiently predictive to enable inference about the parameter of interest. The first-stage estimates are the regression-based analogues to the differences in exposure between treatment and control households seen in Table 5.1.

The first-stage estimates in Table 7.1 are strong for all outcome groups. In all cases, the Kleinbergen-Paap F-statistic exceeds the Staiger and Stock (1997) rule of thumb of 10 for rejecting that instruments are weak. Additionally, they always exceed the Stock and Yogo (2005) critical values for ensuring 2SLS tests have a maximal size smaller than 10% for a Wald test of the coefficient of interest at the 5% level. We are therefore not concerned about bias from a weak first stage affecting our inference about the LATE parameters of interest in any meaningful way.

[18] We use this measure instead of the self-reported information on receipt of the mNutrition messaging because of concerns that households are 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 shown in Section 5 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.

**Table 7.1: First-stage estimates for primary and secondary outcome samples**

| | N | First-stage estimate, basic controls | Kleinbergen-Paap Wald rk F-statistic |
|---|---|---|---|
| Anthropometry | 2,805 | 0.392*** | 232.93 |
| | | (0.026) | |
| Anderson Index: IYCF practices (children aged 6–23 months) | 1,063 | 0.354*** | 98.36 |
| | | (0.036) | |
| Anderson Index: IYCF practices (children aged 6–35 months) | 2,814 | 0.390*** | 235.09 |
| | | (0.025) | |
| WDDS | 2,535 | 0.398*** | 269.28 |
| | | (0.024) | |
| HDDS | 2,585 | 0.396*** | 269.30 |
| | | (0.024) | |
| Female IYCF knowledge | 2,469 | 0.408*** | 296.18 |
| | | (0.024) | |
| Male IYCF knowledge | 2,204 | 0.405*** | 171.31 |
| | | (0.031) | |

Notes: Estimates from the mNutrition Tanzania endline survey sample. Standard errors are in parentheses and clustered at the village level. Reported are the first-stage estimates from the instrumental variable regression, where the mNutrition treatment variable (the indicator for whether the household resided in a village assigned to the treatment group) is used as an excluded instrument for the indicator variable for household-level automatic nutrition message receipt over the last two years, controlling for whether child is the focus child, child age, child gender, and value of the respective outcome at baseline.

Source: Authors' own

## 7.2 Child anthropometry

Table 7.2 shows the LATE estimates of the impact of receipt of nutrition messages on the child anthropometry outcomes. As with the ITT estimates, we are never 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 presented in Table 7.1, implying that they are between two and three times the size of the ITT treatment effects.

**Table 7.2: LATE of impact of mNutrition receipt on child anthropometry, ANCOVA**

| | Control mean | LATE estimates, basic controls | LATE estimates, extended controls | N |
|---|---|---|---|---|
| Child height (cm) | 79.593 | -0.363 | -0.165 | 2,767 |
| | | (0.542) | (0.488) | |
| Child weight (kg) | 10.549 | 0.001 | 0.004 | 2,767 |
| | | (0.159) | (0.161) | |
| Child HAZ | -1.976 | -0.147 | -0.094 | 2,758 |
| | | (0.143) | (0.133) | |
| Child WAZ | -0.933 | -0.034 | -0.043 | 2,766 |
| | | (0.113) | (0.112) | |
| Child WHZ | 0.182 | 0.091 | 0.033 | 2,756 |

| | | | | |
|---|---|---|---|---|
| | | (0.131) | (0.115) | |
| Anderson Index: HAZ and WHZ combined | 0.006 | -0.014 | -0.019 | 2,764 |
| | | (0.080) | (0.079) | |
| Child stunted | 0.501 | 0.074 | 0.060 | 2,758 |
| | | (0.053) | (0.051) | |
| Child underweight | 0.167 | -0.044 | -0.043 | 2,766 |
| | | (0.040) | (0.040) | |
| Child wasted | 0.015 | 0.025 | 0.026 | 2,756 |
| | | (0.015) | (0.016) | |
| Child overweight or obese | 0.036 | 0.013 | 0.009 | 2,756 |
| | | (0.018) | (0.017) | |

Notes: Estimates from the mNutrition Tanzania endline survey sample. Standard errors are in parentheses and clustered at the village level. Reported are the second-stage estimates from the instrumental variable regression, where an indicator variable of household-level mNutrition use over the last two years is instrumented by the treatment variable, controlling for baseline household classification, whether child is the focus child, child's age, child's gender, and value of the respective outcome at baseline. The latter is replaced with 0 for children who were not measured at baseline; indicators for missing measurements are included as controls. 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, and PPI score. Control mean is comparison group's mean at endline. * p<0.10 **p<0.05 ***p<0.01

Source: Authors' own

## 7.3 IYCF practices

Table 7.3 follows the same procedure to produce the LATE estimates of the impact of receipt of mNutrition messaging on IYCF practices for compliers. mNutrition impacts remain statistically significant for the same set of IYCF practices as in the corresponding ITT table, i.e. continued breastfeeding at one year, children's dietary diversity for children aged 6–35 months, the 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, or nearly 50% of the control group mean, although note that 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 from by 0.263, 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 (15.7% of the control group mean) more likely to satisfy the MDD standard. 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 met the MAD by 17.9 percentage points in the basic model (p-value 0.043) and 21.0 percentage points in the extended model (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 in complier households.

There is no statistically significant impact of mNutrition for compliers for either of the two IYCF practices Anderson indices: 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).

**Table 7.3: LATE of impact of mNutrition receipt on IYCF practices, simple differences**

| | Control mean | LATE estimates, basic controls | LATE estimates, extended controls | N |
|---|---|---|---|---|
| Children born in last 24 months who were put to the breast within one hour | 0.776 | 0.012 | 0.018 | 1,001 |
| | | (0.069) | (0.071) | |
| Infants aged 0–5 months who are fed exclusively with breast milk | 0.032 | -0.067 | -0.061 | 165 |
| | | (0.127) | (0.089) | |
| Children aged 12–15 months who are fed breast milk | 0.846 | 0.404** | 0.480*** | 27 |
| | | (0.162) | (0.150) | |
| Infants aged 6–8 months who receive solid, semi-solid, or soft foods | 0.697 | -0.172 | 0.173 | 58 |
| | | (0.439) | (0.515) | |
| Number of food groups (of seven) children aged 6–23 months consume | 3.685 | 0.157 | 0.155 | 833 |
| | | (0.245) | (0.243) | |
| Number of food groups (of seven) children aged 6–35 months consume | 3.766 | 0.263* | 0.280** | 2,524 |
| | | (0.148) | (0.142) | |
| Child consumed grains, roots, or tubers | 0.914 | -0.005 | -0.003 | 2,777 |
| | | (0.021) | (0.022) | |
| Child consumed legumes or nuts | 0.496 | 0.023 | 0.033 | 2,775 |
| | | (0.064) | (0.063) | |
| Child consumed dairy | 0.198 | 0.077 | 0.075 | 2,777 |
| | | (0.052) | (0.051) | |
| Child consumed fish or meat | 0.442 | -0.012 | 0.002 | 2,775 |
| | | (0.060) | (0.058) | |
| Child consumed eggs | 0.080 | 0.014 | 0.015 | 2,775 |
| | | (0.029) | (0.029) | |
| Child consumed vitamin A-rich fruits or vegetables | 0.849 | 0.049 | 0.051* | 2,777 |
| | | (0.032) | (0.031) | |
| Child consumed other fruits or vegetables | 0.579 | 0.093* | 0.087 | 2,775 |
| | | (0.056) | (0.056) | |
| Children aged 6–23 months who consume four or more food groups | 0.609 | 0.024 | 0.015 | 833 |
| | | (0.090) | (0.091) | |
| Children aged 6–35 months who consume four or more food groups | 0.604 | 0.095* | 0.097* | 2,524 |
| | | (0.056) | (0.054) | |
| Children aged 6–23 months who meet the minimum meal frequency | 0.790 | 0.069 | 0.083 | 836 |
| | | (0.069) | (0.069) | |
| Children aged 6–23 months who meet the MAD | 0.224 | 0.179** | 0.210** | 826 |
| | | (0.088) | (0.090) | |
| | -0.003 | 0.021 | 0.028 | 1,001 |

| | | | | |
|---|---|---|---|---|
| Anderson Index: IYCF practices (children aged 6–23 months) | | (0.019) | (0.021) | |
| Anderson Index: IYCF practices (children aged 6–35 months) | -0.001 | 0.008 | 0.009 | 2,700 |
| | | (0.006) | (0.007) | |

Notes: Estimates from the mNutrition Tanzania endline survey sample. Standard errors are in parentheses and clustered at the village level. Reported are the second-stage estimates from the instrumental variable regression, where an indicator variable of household-level mNutrition use over the last two years is instrumented by the 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, and PPI score. Control mean is comparison group's mean at endline. * p<0.10 **p<0.05 ***p<0.01

Source: Authors' own

## 7.4 Dietary diversity

We next turn to estimating the LATE for compliers for WDDS, MDD-W, and HDDS, for which the ITT estimates are shown in Table 6.3. As was the case for the other LATE tables, the LATE estimates in Table 7.4 are between two and three times as large as the ITT estimates, indicating that the treatment effect for households who 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 less precise than the ITT analogues, 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 we 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 (p-value 0.098), while the extended model suggests it increased MDD-W by 7.9 percentage points (p-value 0.125). Both point estimates are large relative to the control group mean, at 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 6.3: the improvements in children's and women's diets were not driven by broader changes in household diets. Instead, the results seen in Table 6.3 and Table 7.4 are more likely due to changes in the way food was allocated across household members.

**Table 7.4: LATE of impact of mNutrition receipt on dietary diversity, ANCOVA**

| | Control mean | LATE estimates, basic controls | LATE estimates, extended controls | N |
|---|---|---|---|---|
| WDDS (0–10) | 4.437 | 0.155 | 0.164 | 2,510 |
| | | (0.168) | (0.165) | |
| Met MDD-W | 0.625 | 0.088* | 0.079 | 2,521 |
| | | (0.053) | (0.051) | |
| HDDS (0–12) | 5.336 | 0.037 | 0.099 | 2,549 |
| | | (0.224) | (0.211) | |

Notes: Estimates from the mNutrition Tanzania endline survey sample. Standard errors are in parentheses and clustered at the village level. Reported are the second-stage estimates from the instrumental variable regression, where an indicator variable of household-level mNutrition use over the last two years is instrumented by the 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, and PPI score. Control mean is comparison group's mean at endline. * p<0.10 **p<0.05 ***p<0.01

Source: Authors' own

## 7.5 IYCF knowledge and beliefs

Finally, Table 7.5 shows the mNutrition LATE estimates for household compliers for the IYCF knowledge outcomes. The point estimate for females now indicates that primary females in households induced to receive the mNutrition messages by the random offer of treatment answered 1.95% more questions correctly in the basic model (p-value 0.222) and 1.841% more questions correctly in the extended model (p-value 0.226), though neither estimate is statistically significant at the 10% level.

The LATE estimate for male knowledge is more than twice as large as the female estimate and statistically significant at the 5% level in both the basic and the extended models. Men in complier households answered 4.407% more questions correctly in the basic model (p-value 0.018) and 4.343% more questions correctly in the extended model (p-value 0.021) as a result of the mNutrition content. The point estimates imply that males in complier households answered nearly 0.9 additional questions correctly (out of 20).

The household-level Anderson Index for IYCF knowledge is also statistically significantly higher in complier households induced to receive the mNutrition content: receipt of the messages increased IYCF knowledge by 0.23 standard deviations in the basic model and 0.232 standard deviations in the extended model, both of which are significant at the 5% level. This implies roughly a 2.92 percentage point increase in the number of questions answered correctly when averaging across the male and female scores. Clearly, the mNutrition service importantly increased household knowledge of nutrition and IYCF practices, with males being the primary recipients of the new knowledge.

**Table 7.5: LATE of impact of mNutrition receipt on females' IYCF knowledge, ANCOVA**

| | Control mean | LATE estimates, basic controls | LATE estimates, extended controls | N |
|---|---|---|---|---|
| Percent of answers correct (female) | 68.969 | 1.950 | 1.841 | 2,469 |
| | | (1.597) | (1.520) | |
| Percent of answers correct (male) | 58.959 | 4.407** | 4.343** | 2,204 |
| | | (1.863) | (1.887) | |
| Anderson Index: combined percent correct of females and males | -0.047 | 0.230** | 0.232** | 2,114 |
| | | (0.105) | (0.100) | |

Notes: Estimates from the mNutrition Tanzania endline survey sample. Standard errors are in parentheses and clustered at the village level. Reported are the second-stage estimates from the instrumental variable regression, where an indicator variable of household-level mNutrition use over the last two years is instrumented by the 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, and PPI score. Control mean is comparison group's mean at endline. * p<0.10 **p<0.05 ***p<0.01

Source: Authors' own

## 7.6 Summary

The results in Section 7 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, though the LATE estimates presented in Section 7 are always less precise than the corresponding ITT estimates, and therefore only statistically distinguishable from zero for seven of the anthropometric, IYCF practices, dietary diversity, and IYCF knowledge outcomes. 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 presented in Section 6.

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, we 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 we 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 being the measure of receipt of the mNutrition content, we 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 our 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 2017 completely identifies the treatment households who 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 in the tables presented in this section. 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, as shown in Table 7.1 to Table 7.5, an important question is whether the LATE estimates or the ITT estimates are more relevant for policymakers when evaluating the usefulness of the mNutrition service. Although the LATE estimates provide critical information about the potential benefits of the information contained within the mNutrition content, acknowledging that service take-up is imperfect even among households and individuals who express a willingness to participate is important when calculating the benefits of the service. 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 frictions in relation to disseminating information through mobile phones should be incorporated when comparing the effectiveness of mobile phone-based interventions to other methods of BCC. We therefore prefer the ITT estimates presented in Section 6 to the LATE estimates shown in this section for characterising the service benefits. However, exploring ways to decrease the barriers to sharing nutrition information with households in developing countries through mobile phones is an interesting area for subsequent research. This is especially true as the degree of attachment to mobile phones changes with the prevalence of smartphones and as the availability of mobile internet changes how individuals interact with their mobile phones in developing countries.

# 8 Sub-randomisation and heterogeneity

The ITT estimates presented in Section 6 capture the average impact of the offer of access to the mNutrition messaging for the study population. How the service impact differed for distinct sub-populations in the overall study sample is also of interest and is a question that can be answered given the data we collected and the evaluation design. In this section, therefore, we focus on two distinct sources of treatment effect heterogeneity.

First, we use the randomly assigned variation in whether just the primary female or both the primary female and the primary male were offered access to the mNutrition messaging in households where the primary female and primary male owned different mobile phones, to explore whether sending the content to both household members leads to any differential impact of the service on the primary or secondary outcomes. Using the same variation, we are also able to estimate the mNutrition service impacts when we intentionally select the primary female as the mNutrition recipient. As noted in the baseline report (Gilligan *et al.,* 2018), the households eligible for this sub-randomisation represent only a small subset of the overall sample: of the 1,428 treatment households interviewed at baseline, just 276 (19.3%) were identified as being eligible for the household-level randomisation. Relative to the full baseline sample, eligible households had higher education levels, were less likely to be female-headed, had higher PPI scores, and higher scores on the different asset indices constructed using the baseline data. We therefore have limited statistical power to make inferences about these secondary analyses, and results for this sub-sample may not be generalisable to the households in the broader sample.

Second, we also explore whether there is any heterogeneity in the mNutrition service impacts by baseline household wealth. To do so, we use the PPI scores for each household to map them to the likelihood that they fall below 150% of the national poverty line (NPL) in Tanzania. We then interact this probability with the mNutrition treatment indicator to allow the impact of mNutrition to differ along this dimension. We similarly interact the other controls with the treatment indicator so that the main (non-interacted) coefficient on mNutrition treatment represents the overall average treatment effect of the service, and the interaction between the poverty likelihood and the treatment indicator measures how this impact changes with the baseline poverty likelihood of the household.[19]

For both the sub-randomisation and heterogeneity by baseline household poverty, the tables in the sub-sections below show the estimates for the primary and secondary outcomes and their respective components. We include all households from either treatment group where the primary female and primary male were both interviewed at baseline and where they reported owning distinct mobile phones. The first column in the table presents the mean value in the endline survey for all control households that would have been eligible for the household-level randomisation while the last column shows the total number of treatment and control observations that were eligible. The second and third columns display the mNutrition impact estimate for being randomly assigned to the female-only sub-treatment arm (T-F) and being assigned to the female + male sub-treatment arm (T-F+M), relative to being from a household eligible for the household-level randomisation but from a control village. We also show the p-value from an F-test of the null hypothesis that there is no difference between the impact of being assigned to the T-F group and the impact of being assigned to the T-F+M group. All estimates and p-values are based on standard errors that are clustered at the village level and are based on the basic model, as discussed in previous sections.

[19] In practice, we follow Imbens and Rubin (2015) and interact the de-meaned poverty likelihood and control variables with the treatment indicator while still including the unadjusted values of the characteristics as controls.

## 8.1 Sub-randomisation

### 8.1.1 Child anthropometry

Table 8.1 begins by displaying the estimated coefficients for anthropometry for the sub-sample eligible for the household-level randomisation. As was the case in both the ITT and LATE tables, we find no evidence of any impact of mNutrition on child anthropometry for the eligible sub-population. The point estimates for HAZ and WAZ are both negative, but small in magnitude and imprecise, while the point estimates for stunting are positive for the T-F group and negative for the T-F+M group. The impact of mNutrition on the Anderson Index that combines the HAZ and WHZ outcomes is similarly small in magnitude, with p-values of 0.865 for the T-F indicator and 0.909 for the T-F+M indicator. We therefore find no evidence that sending messages to the primary male as well as to the primary female had any differential impact on child anthropometry.

**Table 8.1: Heterogeneous impacts of mNutrition on child anthropometry, sub-randomisation arms and eligible control group households**

| | Mean of eligible control group | Female-only treatment household (T-F) | Female and male treatment household (T-F+M) | P-value of T-F=T-F+M | N |
|---|---|---|---|---|---|
| Child height (cm) | 80.597 | -0.082 | -0.470 | 0.464 | 508 |
| | | (0.440) | (0.462) | | |
| Child weight (kg) | 10.785 | 0.024 | -0.030 | 0.705 | 508 |
| | | (0.125) | (0.125) | | |
| Child HAZ | -1.821 | -0.052 | -0.093 | 0.756 | 506 |
| | | (0.117) | (0.113) | | |
| Child WAZ | -0.804 | -0.016 | -0.061 | 0.703 | 508 |
| | | (0.095) | (0.106) | | |
| Child WHZ | 0.207 | 0.029 | 0.041 | 0.919 | 506 |
| | | (0.100) | (0.107) | | |
| Anderson Index: HAZ and WHZ combined | 0.080 | 0.012 | -0.009 | 0.814 | 507 |
| | | (0.069) | (0.074) | | |
| Child stunted | 0.455 | 0.024 | -0.022 | 0.392 | 506 |
| | | (0.049) | (0.051) | | |
| Child underweight | 0.122 | 0.014 | -0.037 | 0.158 | 508 |
| | | (0.034) | (0.031) | | |
| Child wasted | 0.012 | 0.002 | 0.021 | 0.356 | 506 |
| | | (0.011) | (0.019) | | |
| Child overweight or obese | 0.020 | 0.002 | 0.021 | 0.362 | 506 |
| | | (0.015) | (0.019) | | |

Notes: Estimates from the mNutrition Tanzania endline survey sub-sample. Standard errors are in parentheses and clustered at the village level. Reported are the impact estimates of sub-randomisation of eligible households into households where only the female receives mNutrition content, households where both the primary female and primary male receive content, and control households. The eligibility criterion was ownership of a separate mobile phone by both the female and male. Controls include whether child is the focus child, child age, child gender, and value of the respective outcome at baseline. The latter is replaced with 0 for children who were not measured at baseline; indicators for missing measurements are included as controls. * p<0.10 **p<0.05 ***p<0.01

Source: Authors' own

### 8.1.2 IYCF practices

Table 8.2 presents the same parameters for the IYCF practices outcomes. The ITT estimates in Table 6.2 and the LATE estimates in Table 7.2 uncovered impacts of access or receipt of the mNutrition service on the likelihood that children aged 12–15 months were still breastfed, CDDS for children aged 6–35 months, the likelihood of children aged 6–35 months achieving MDD, and the likelihood that children aged 6–23 months satisfy MAD. There are too few observations in the eligible sub-sample to estimate impacts on continued breastfeeding for 12–15-month-old children and we find no statistically significant impacts of being assigned to the T-F group or the T-F+M group on CDDS or MDD for children aged 6–35 months, or MAD for children aged 6–23 months.

While the CDDS point estimate for the T-F group (0.062) is comparable in size to the ITT point estimate in Table 6.2 (0.107), we are unable to reject the possibility that there is no difference between the magnitudes. The T-F+M point estimate for CDDS is negatively signed but sufficiently imprecise that we are likewise unable to reject the possibility that it is equal to the estimate for the T-F group: the p-value from the test of whether the two coefficients are equal is 0.474. The estimates for MDD for children aged 6–35 months and MAD for children aged 6–23 months similarly suggest there are no statistically significant differences between children in eligible control households, children in the T-F sub-treatment group, and children in the T-F+M sub-treatment group. The point estimates are positive for the T-F group and negative for the T-F+M group, but the small sample and effect sizes mean we cannot reject the possibility that the average level of the outcomes is the same for children from all three groups.

The impact estimates for the Anderson indices for IYCF practices for children aged 6–23 months and 6–35 months are both positively signed but, again, statistically insignificantly different from zero and from one another. In the full tables of IYCF practice estimates for the sub-randomisation arms, there is only one coefficient indicating that either of the sub-randomisation treatments had a statistically significant impact on an outcome relative to the children in eligible control households: children in the T-F+M arm were less likely to have consumed fish or meat in the day preceding the survey. There are also only three outcomes for which the estimated impact of being assigned to the T-F arm is statistically significantly different from the impact of being assigned to the T-F+M arm: T-F+M children were less likely than T-F children to have consumed chicken or meat, more likely than T-F children to have consumed grains, roots, or tubers, and more likely than T-F children to have consumed vitamin A-rich fruits or vegetables.

**Table 8.2: Heterogeneous impacts of mNutrition on IYCF practices, sub-randomisation arms and eligible control group households**

| | Mean of eligible control group | Female-only treatment household (T-F) | Female and male treatment household (T-F+M) | P-value of T-F=T-F+M | N |
|---|---|---|---|---|---|
| Children born in last 24 months who were put to the breast within one hour | 0.739 | 0.010 | -0.034 | 0.583 | 193 |
| | | (0.072) | (0.067) | | |
| Infants aged 0–5 months who are fed exclusively with breast milk | 0.000 | 0.000 | 0.000 | | 27 |
| | | (0.000) | (0.000) | | |
| Number of food groups (of seven) | 3.944 | -0.105 | -0.264 | 0.506 | 165 |
| | | (0.231) | (0.212) | | |

| | | | | | |
|---|---|---|---|---|---|
| children aged 6–23 months consume | | | | | |
| Number of food groups (of seven) children aged 6–35 months consume | 3.970 | 0.062 | -0.057 | 0.474 | 469 |
| | | (0.140) | (0.128) | | |
| Child consumed grains, roots, or tubers | 0.929 | -0.032 | 0.027 | 0.015** | 510 |
| | | (0.028) | (0.018) | | |
| Child consumed legumes or nuts | 0.512 | 0.012 | 0.001 | 0.878 | 510 |
| | | (0.057) | (0.061) | | |
| Child consumed dairy | 0.280 | -0.004 | -0.064 | 0.300 | 510 |
| | | (0.054) | (0.051) | | |
| Child consumed fish or meat | 0.520 | 0.023 | -0.121** | 0.034** | 510 |
| | | (0.055) | (0.054) | | |
| Child consumed eggs | 0.075 | 0.030 | 0.017 | 0.727 | 510 |
| | | (0.034) | (0.033) | | |
| Child consumed vitamin A-rich fruits or vegetables | 0.870 | -0.030 | 0.036 | 0.043** | 510 |
| | | (0.036) | (0.027) | | |
| Child consumed other fruits or vegetables | 0.594 | 0.052 | 0.042 | 0.853 | 510 |
| | | (0.050) | (0.052) | | |
| Children aged 6–23 months who consume four or more food groups | 0.722 | -0.088 | -0.128 | 0.684 | 165 |
| | | (0.090) | (0.091) | | |
| Children aged 6–35 months who consume four or more food groups | 0.690 | 0.017 | -0.026 | 0.566 | 469 |
| | | (0.061) | (0.059) | | |
| Children aged 6–23 months who meet the minimum meal frequency | 0.816 | -0.062 | -0.076 | 0.877 | 166 |
| | | (0.067) | (0.068) | | |
| Children aged 6–23 months who meet the MAD | 0.324 | 0.073 | -0.008 | 0.411 | 163 |
| | | (0.090) | (0.095) | | |
| Anderson Index: IYCF practices (children aged 6–23 months) | -0.017 | 0.008 | 0.018 | 0.209 | 193 |
| | | (0.021) | (0.021) | | |
| Anderson Index: IYCF practices (children aged 6–35 months) | -0.006 | 0.003 | 0.008 | 0.110 | 498 |
| | | (0.008) | (0.008) | | |

Notes: Estimates from the mNutrition Tanzania endline survey sub-sample. Standard errors are in parentheses and clustered at the village level. Reported are the impact estimates of sub-randomisation of eligible households into households where only the female receives mNutrition content, households where both the primary female and primary male receive content, and control households. The eligibility criterion was ownership of a separate mobile phone by both the female and male. Model controls for child's age and gender. * p<0.10 **p<0.05 ***p<0.01

Source: Authors' own

### 8.1.3 Dietary diversity

Table 8.3 shows the mNutrition impact estimates for the T-F and T-F+M sub-randomisation arms for the dietary diversity outcomes. WDDS scores are 0.359 higher, on average, for women in the T-F sub-treatment arm relative to women in eligible control households (p-value 0.012). This effect size is equivalent to 7.8% of the control group mean WDDS among households eligible for the sub-randomisation. We are also able to reject the possibility that there is no difference between the point estimate for the T-F group and the T-F+M group, with the estimate for the T-F+M being statistically significantly lower than the estimate for the T-F.

Consistent with the WDDS results, T-F women are also more likely to satisfy MDD-W. The point estimate indicates that women in treatment households assigned to the T-F group are 9.4 percentage points more likely to satisfy MDD-W than their control group counterparts. This is equivalent to 14.1% of the control group mean. T-F+M women are statistically significantly less likely than T-F women to achieve MDD-W (p-value 0.082) but not differentially likely to satisfy MDD-W than eligible control group women. HDDS is statistically significantly lower in the T-F+M group than in either the T-F group or among the eligible control households. The point estimate suggests T-F+M households consume from 0.473 fewer categories than eligible control households, or 8.3% of the control group mean.

The results in Table 8.3 therefore provide some evidence that women in T-F households may consume more diverse diets than women in eligible control households or women in T-F+M households. Furthermore, T-F+M women do not have higher WDDS and are not more likely to satisfy MDD-W than women in eligible control households, suggesting that there is no advantage for women's diets to sending text messages to the women and their spouses, even relative to not sending text messages to any household member. Given the smaller sample size, we are reluctant to place too much weight on these results. To the extent that these impacts are not just driven by statistical chance and small sample variability, there are two mechanisms that could explain these findings. First, sending duplicate messages to multiple household members could overburden households, and lead to recipients being less likely to open, read, or remain subscribed to the service. Pop-Eleches *et al*. (2011) similarly find that individuals randomly assigned to receive SMS reminders weekly were more likely than individuals who did not receive any reminders to adhere to an antiretroviral therapy routine, but individuals who received daily reminders were no more likely to adhere to the prescribed treatment. Second, sending duplicate messages to two household members could generate a type of 'free rider' problem, where each spouse becomes less likely to open, read, or adhere to the information contained in a message because they prefer to not incur any costs associated with these actions in the hope that their spouse will undertake these actions.

**Table 8.3: Heterogeneous impacts of mNutrition on dietary diversity, sub-randomisation arms and eligible control group households**

| | Mean of eligible control group | Female-only treatment household (T-F) | Female and male treatment household (T-F+M) | P-value of T-F=T-F+M | N |
|---|---|---|---|---|---|
| WDDS (0–10) | 4.623 | 0.359** | -0.123 | 0.002*** | 466 |
| | | (0.142) | (0.141) | | |
| Met MDD-W | 0.667 | 0.094* | -0.010 | 0.082* | 467 |
| | | (0.049) | (0.053) | | |
| HDDS (0–12) | 5.696 | 0.059 | -0.473** | 0.006*** | 473 |
| | | (0.182) | (0.186) | | |

Notes: Estimates from the mNutrition Tanzania endline survey sub-sample. Standard errors are in parentheses and clustered at the village level. Reported are the impact estimates of sub-randomisation of eligible households into households where only the female receives mNutrition content, households where both the primary female and primary male receive content, and control households. The eligibility criterion was ownership of a separate mobile phone by both the female and male. Controls include household type at baseline (pregnant mother or child younger than 12 months) and the respective outcome at baseline. * p<0.10 **p<0.05 ***p<0.01

Source: Authors' own

### 8.1.4 IYCF knowledge and beliefs

Lastly, Table 8.4 shows the impact estimates of the impact of being assigned to the T-F and T-F+M groups on IYCF knowledge, relative to being in an eligible control household. We find no statistically significant differences in the percentage of correct answers for women, men, or for the Anderson Index of household-level IYCF knowledge. Similarly, we find no differences between the T-F and T-F+M groups, with 0.574 being the smallest p-value from the F-test of whether there is no difference between the point estimates.

The point estimates for female knowledge are negatively signed for both the T-F and T-F+M groups, but the t-statistics are always less than one. The male point estimates are positive for both T-F and T-F+M, with corresponding p-values of 0.147 and 0.102. If we were to combine the T-F and T-F+M indicators to calculate the impact on male IYCF knowledge of being assigned to the mNutrition treatment group for households where the primary female and primary male owned distinct mobile phones, we would estimate a coefficient on mNutrition treatment of 2.327 (p-value 0.061). Thus, after adjusting the estimation strategy to more closely mimic the ITT models in Section 6, at least the IYCF knowledge result for males persists in the sub-sample of households that were eligible for the household-level randomisation.

Impacts on the household-level IYCF knowledge Anderson Index are positively signed but not statistically significantly different from zero for either the T-F or T-F+M group, with both t-statistics being below one. The T-F and T-F+M coefficients for combined IYCF knowledge are not different from one another, again suggesting there is no additional benefit from sending content to the mobile phone of both the primary female and the primary male.

**Table 8.4: Heterogeneous impacts of mNutrition on IYCF knowledge, sub-randomisation arms and eligible control group households**

| | Mean of eligible control group | Female-only treatment household (T-F) | Female and male treatment household (T-F+M) | P-value of T-F=T-F+M | N |
|---|---|---|---|---|---|
| Percent of answers correct (female) | 70.982 | -0.978 | -0.122 | 0.574 | 457 |
| | | (1.315) | (1.337) | | |
| Percent of answers correct (male) | 61.560 | 2.223 | 2.420 | 0.908 | 448 |
| | | (1.525) | (1.471) | | |
| Anderson Index: combined percent correct of females and males | 0.097 | 0.047 | 0.068 | 0.805 | 431 |
| | | (0.078) | (0.078) | | |

Notes: Estimates from the mNutrition Tanzania endline survey sub-sample. Standard errors are in parentheses and clustered at the village level. Reported are the impact estimates of sub-randomisation of eligible households into households where only the female receives mNutrition content, households where both the primary female and primary male receive content, and control households. The eligibility criterion was ownership of a separate mobile phone by both the female and male. Controls include household type at baseline (pregnant mother or child younger than 12 months) and the respective outcome at baseline. * p<0.10 **p<0.05 ***p<0.01

Source: Authors' own

Though it is surprising that we do not find statistically significant impacts of the mNutrition service on any of the IYCF knowledge indicators, this is likely to be primarily driven by two factors. First, as shown in the baseline report (Gilligan *et al.,* 2018), the sample of households eligible for the sub-randomisation is different from the overall sample along a number of dimensions. At endline, control households eligible for the sub-randomisation have females that answered 2.0% more IYCF questions correctly and males that answered 2.6% more IYCF questions correctly than the average female and male from across all control households. Thus, eligible households have higher levels of IYCF knowledge than non-eligible households in the sample. This additional knowledge may make it more difficult for the mNutrition content to improve individual knowledge levels as it increases the likelihood that the information is not new for the service beneficiaries. Second, as mentioned above, the sample size is less than one-fifth the size for this sub-randomisation analysis. Starting with this smaller sample, we then further reduce the statistical power we have to detect differences by splitting the full treatment group into the T-F and T-F+M indicators. Both features of the sub-randomisation analysis therefore reduce the likelihood of our finding impacts on IYCF knowledge, or any of the primary and secondary outcomes more generally.

What is less ambiguous from Table 8.1 to Table 8.4 is that we do not find evidence that sending the mNutrition messages to the mobile phone of the primary male in addition to the primary female has a differential impact on any of the primary or secondary outcomes. In fact, we more often reject the null of no difference between the estimate for the T-F and T-F+M groups because the service was more effective for the T-F group (four outcomes: the three dietary diversity measures and children's consumption of fish or meat) than because the impact was larger for the T-F+M group (two outcomes: children's consumption of grains, roots, or tubers and children's consumption of vitamin A-rich fruits and vegetables).

## 8.2 Heterogeneity in mNutrition impacts by household wealth

In this section we turn to exploring whether the impact of access to the mNutrition messaging differed by household wealth at baseline. We use the PPI to map household characteristics and asset ownership to a likelihood that households were below 150% of the NPL in Tanzania.[20] For the households in our sample these likelihoods range from 0 to 100, with a mean likelihood of 56.1% and a median likelihood of 58.1%.

The impact of access to mNutrition could differ by household wealth for a multitude of reasons. For example, wealthier households may be more likely to have access to other reliable sources for nutrition information, such as government health workers, health facilities, or more knowledgeable peers. The mNutrition information may be more likely to duplicate existing knowledge for individuals from wealthier households, suggesting that we should expect to find larger impacts of access to the service on the primary and secondary outcomes for less wealthy households. Conversely, individuals in wealthier households may have higher literacy rates, or greater ability to understand the message content, or it could be that implementing the suggestions in the mNutrition content requires household resources: for example, purchasing animal-sourced foods may be costly and therefore less common for poor households, even conditional on knowing the potential benefits of them for young children. This would suggest we might expect to find larger impacts of access to mNutrition for wealthier households. Thus, we have no strong theoretical reason to expect the relationship between household wealth and the mNutrition service impacts to be positive or negative and we turn to estimating this empirically.

[20] For more details on the PPI we refer interested readers to Annex H of the baseline report for this project (Gilligan *et al*., 2018) or to the PPI website: www.progressoutofpoverty.org/.

### 8.2.1 Child anthropometry

Table 8.5 begins by displaying the coefficients for the randomly assigned offer of access to the mNutrition service indicator, the likelihood that the household was below 150% of the NPL based on their baseline characteristics, and the interaction between the mNutrition service indicator and the likelihood that the household was below 150% of the NPL at baseline after de-meaning using the overall sample mean. The coefficient on the interaction can therefore be interpreted as the additional impact of the mNutrition service for households that were identified as being one percentage point more likely to be below 150% of the NPL relative to a household at the sample mean likelihood.

The (non-causal) associations between the household likelihood of being below 150% of the NPL are typically of the expected sign, suggesting children from less wealthy households have lower HAZ, WHZ, and WAZ scores, and lower values of the combined HAZ/WHZ Anderson Index. We find statistically significant, albeit small, positive associations between household wealth and the likelihood that children are wasted or underweight.

The mNutrition treatment effects at the sample mean likelihood of being below 150% NPL are close to the estimates without any interactions we presented in Table 6.1: we never find a statistically significant effect and the point estimates are always small in size. The treatment effects for the z-scores, for example, are all below 0.06 in absolute value, re-affirming the main estimates that suggested there was no impact of access to the mNutrition service on child anthropometry.

The treatment interactions are always signed to suggest that access to the mNutrition service had a more beneficial impact for children from households that were deemed more likely to be below 150% of the NPL at baseline. That said, all the point estimates are small in magnitude and only one is statistically significantly different from zero (for the likelihood that children are wasted), offering weak evidence that the service may have been more beneficial for less wealthy households.

**Table 8.5: Heterogeneity in mNutrition impacts on child anthropometry by household wealth**

| | Treatment | Likelihood of <150% NPL | Treatment x likelihood of <150% NPL | N |
|---|---|---|---|---|
| Child height (cm) | -0.148 | -0.008* | 0.001 | 2,803 |
| | (0.210) | (0.005) | (0.007) | |
| Child weight (kg) | -0.007 | -0.004** | 0.002 | 2,803 |
| | (0.062) | (0.002) | (0.002) | |
| Child HAZ | -0.059 | -0.002 | 0.000 | 2,794 |
| | (0.055) | (0.001) | (0.002) | |
| Child WAZ | -0.017 | -0.003*** | 0.001 | 2,802 |
| | (0.043) | (0.001) | (0.002) | |
| Child WHZ | 0.030 | -0.003** | 0.001 | 2,792 |
| | (0.051) | (0.001) | (0.002) | |
| Anderson Index: HAZ and WHZ combined | -0.009 | -0.002** | 0.001 | 2,800 |
| | (0.030) | (0.001) | (0.001) | |
| Child stunted | 0.028 | 0.000 | -0.000 | 2,794 |
| | (0.020) | (0.000) | (0.001) | |
| Child underweight | -0.015 | 0.001* | -0.001 | 2,802 |

| | | | | |
|---|---|---|---|---|
| | (0.015) | (0.000) | (0.001) | |
| Child wasted | 0.010 | 0.000** | -0.001* | 2,792 |
| | (0.006) | (0.000) | (0.000) | |
| Child obese | 0.004 | -0.000 | 0.000 | 2,792 |
| | (0.007) | (0.000) | (0.000) | |

Notes: Estimates from the mNutrition Tanzania endline survey sample. Standard errors are in parentheses and clustered at the village level. Reported are impact estimates from the regressions of each outcome on the treatment indicator, likelihood of being below 150% of the NPL, and an interaction term of the latter two (the estimand of interest, which is presented in column 3). The poverty likelihood is converted from the PPI using the threshold values created for Tanzania. The model controls for whether child is the focus child, child age, child gender, and value of the respective outcome at baseline. The latter is replaced with 0 for children who were not measured at baseline; indicators for missing measurements are included as controls. * p<0.10 **p<0.05 ***p<0.01

Source: Authors' own

### 8.2.2 IYCF practices

We next turn to exploring heterogeneity in the mNutrition impacts on IYCF practices by household wealth. In the main estimates, we found evidence that the mNutrition service significantly improved the likelihood that children aged 12–15 months were still breastfed, the likelihood that children consumed vitamin A-rich fruits or vegetables as well as other fruits and vegetables, the number of food groups from which children aged 6–35 months consumed during the day preceding the survey, the likelihood that children aged 6–35 months satisfied MDD, and the likelihood that children aged 6–23 months satisfied MAD.

As with the child anthropometry outcomes, the household wealth associations with the IYCF practices outcomes are typically signed to indicate that wealthier households are more likely to comply with IYCF recommendations. Children in wealthier households are more likely to be fed solid, semi-solid, or soft foods when aged between 6–8 months, they consume from more food groups between both 6–23 and 6–35 months (with legumes or nuts, fish or meat, eggs, vitamin A-rich fruits or vegetables, and other fruits and vegetables the most affected food categories), they are more likely to satisfy MDD between 6–23 months and 6–35 months, and they are more likely to satisfy minimum meal frequency and MAD when aged between 6–23 months.

The mNutrition treatment effects at the sample means of the control variables and the likelihood of the household being below 150% of the NPL are also nearly identical to the main estimates shown in Table 6.2. Table 8.6 suggests that mNutrition had statistically significant impacts on the likelihood that children aged 12–15 months were still breastfed,[21] CDDS for children aged 6–35 months (which is driven primarily by increases in the likelihood children consumed vitamin A-rich fruits or vegetables and other fruits or vegetables), an increased likelihood that children aged 6–35 months satisfy MDD, and an increased likelihood that children aged 6–23 months satisfy MAD.

Turning to the estimates of how the mNutrition impact varies with baseline household wealth, Table 8.6 offers more evidence that the service was more beneficial for less wealthy households. In all cases where the point estimates are not zero (to at least three decimal places), the sign of the interaction term indicates that the impact was larger for less wealthy households. For the number of food groups consumed (children aged 6–23 months), and the likelihood that children satisfy MDD (for children aged 6–23 months and 6–35 months), the interaction terms are statistically significantly different from zero and point toward meaningfully sized heterogeneity in the impacts. For example, for a child aged 6–23 months in a household at the mean probability of being below the NPL, access to the mNutrition service increased CDDS by .065, although the

[21] We emphasise again the previously mentioned caveats about the extremely small sample size for this outcome.

estimate is not statistically significantly different from zero. For a child from a household that was 20 percentage points more likely to be below 150% of the NPL—just below a one standard deviation increase—access to the mNutrition service increases CDDS by 0.190 (p-value 0.071).

**Table 8.6: Heterogeneity in mNutrition impacts on IYCF practices by household wealth**

| | Treatment | Likelihood of <150% NPL | Treatment x likelihood of <150% NPL | N |
|---|---|---|---|---|
| Children born in last 24 months who were put to the breast within one hour | 0.009 | 0.000 | 0.000 | 1,011 |
| | (0.025) | (0.001) | (0.001) | |
| Infants aged 0–5 months who are fed exclusively with breast milk | -0.013 | -0.000 | 0.000 | 165 |
| | (0.027) | (0.001) | (0.001) | |
| Children aged 12–15 months who are fed breast milk | 0.284*** | -0.002 | 0.002 | 27 |
| | (0.097) | (0.003) | (0.003) | |
| Infants aged 6–8 months who receive solid, semi-solid, or soft foods | -0.007 | -0.008** | 0.003 | 58 |
| | (0.136) | (0.004) | (0.006) | |
| Number of food groups (of seven) children aged 6–23 months consume | 0.065 | -0.010*** | 0.006* | 843 |
| | (0.089) | (0.003) | (0.003) | |
| Number of food groups (of seven) children aged 6–35 months consume | 0.115** | -0.008*** | 0.003 | 2,558 |
| | (0.056) | (0.001) | (0.002) | |
| Child consumed grains, roots, or tubers | -0.001 | -0.000** | 0.001** | 2,813 |
| | (0.008) | (0.000) | (0.000) | |
| Child consumed legumes or nuts | 0.010 | -0.002*** | 0.001 | 2,811 |
| | (0.025) | (0.001) | (0.001) | |
| Child consumed dairy | 0.031 | -0.001 | -0.001 | 2,813 |
| | (0.020) | (0.001) | (0.001) | |
| Child consumed fish or meat | -0.000 | -0.001** | 0.000 | 2,811 |
| | (0.023) | (0.001) | (0.001) | |
| Child consumed eggs | 0.006 | -0.001** | 0.000 | 2,811 |
| | (0.011) | (0.000) | (0.001) | |
| Child consumed vitamin A-rich fruits or vegetables | 0.022* | -0.001* | 0.001 | 2,813 |
| | (0.012) | (0.000) | (0.000) | |
| Child consumed other fruits or vegetables | 0.041* | -0.002*** | 0.000 | 2,811 |
| | (0.022) | (0.001) | (0.001) | |
| Children aged 6–23 months who consume four or more food groups | 0.004 | -0.003*** | 0.002* | 843 |
| | (0.034) | (0.001) | (0.001) | |
| Children aged 6–35 months who consume four or more food groups | 0.041* | -0.003*** | 0.001* | 2,558 |
| | (0.021) | (0.001) | (0.001) | |
| Children aged 6–23 months who meet the minimum meal frequency | 0.026 | -0.002** | 0.001 | 846 |
| | (0.026) | (0.001) | (0.001) | |
| Children aged 6–23 months who meet the MAD | 0.073** | -0.002** | -0.000 | 830 |
| | (0.034) | (0.001) | (0.001) | |
| Anderson Index: IYCF practices (children aged 6–23 months) | 0.008 | -0.000 | 0.000 | 1,011 |
| | (0.007) | (0.000) | (0.000) | |
| | 0.003 | -0.000 | 0.000 | 2,734 |

| | | | | |
|---|---|---|---|---|
| Anderson Index: IYCF practices (children aged 6–35 months) | (0.002) | (0.000) | (0.000) | |

Notes: Estimates from the mNutrition Tanzania endline survey sample. Standard errors are in parentheses and clustered at the village level. Reported are impact estimates from the regressions of each outcome on the treatment indicator, likelihood of being below 150% of the NPL, and an interaction term of the latter two. The poverty likelihood is converted from the PPI using the threshold values created for Tanzania. The model controls for child's age and gender. * p<0.10 **p<0.05 ***p<0.01

Source: Authors' own

### 8.2.3 Dietary diversity

Table 8.7 displays the results of the same household wealth heterogeneity specifications for the dietary diversity outcomes. The associations between household wealth and dietary diversity for women and households are among the strongest seen so far in this section: all three have p-values below 0.01 and suggest that diets for women and household members more broadly are importantly more diverse in wealthier households. The non-interacted treatment estimates indicate that women in mNutrition service households have more diverse diets: with higher WDDS and an increased likelihood of meeting MDD-W, although only the latter result is statistically different from zero. The HDDS treatment effect is positive but small and not distinguishable from zero, as was the case in Table 6.3.

The interaction terms, again, all suggest that access to the service had a larger impact on less wealthy households. However, the treatment effect estimates all have t-statistics around 1, and we can therefore never reject the possibility that there is no wealth gradient in the mNutrition service impacts on dietary diversity.

**Table 8.7: Heterogeneity in mNutrition impacts on dietary diversity by household wealth**

| | Treatment | Likelihood of <150% NPL | Treatment x likelihood of <150% NPL | N |
|---|---|---|---|---|
| WDDS (0–10) | 0.080 | -0.007*** | 0.003 | 2,535 |
| | (0.066) | (0.002) | (0.003) | |
| Met MDD-W | 0.042** | -0.002*** | 0.001 | 2,546 |
| | (0.021) | (0.001) | (0.001) | |
| HDDS (0–12) | 0.025 | -0.010*** | 0.004 | 2,585 |
| | (0.086) | (0.002) | (0.003) | |

Notes: Estimates from the mNutrition Tanzania endline survey sample. Standard errors are in parentheses and clustered at the village level. Reported are impact estimates from the regressions of each outcome on the treatment indicator, likelihood of being below 150% of the NPL, and an interaction term of the latter two. The poverty likelihood is converted from the PPI using the threshold values created for Tanzania. The model controls for the value of the respective outcome at baseline. * p<0.10 **p<0.05 ***p<0.01

Source: Authors' own

### 8.2.4 IYCF knowledge and beliefs

Last, we turn to exploring whether there is any heterogeneity by household wealth in the mNutrition service impacts on IYCF knowledge and beliefs. Household wealth is strongly predictive of IYCF knowledge, with wealthier households possessing more information and more accurate beliefs about IYCF practices. A female from a household that is 20 percentage points less likely to be below the NPL answers 1.34% more questions correctly, while a male answers 1.78% more questions correctly on average. These amount to differences equivalent to 1.94% and 3.02% of the respective control group means from a change in household wealth slightly smaller than one standard deviation. The non-interacted treatment effect estimates are nearly indistinguishable from the main ITT estimates, suggesting there are statistically significant increases in male IYCF knowledge and in the household-level aggregate, and a positive but statistically insignificant increase in female IYCF knowledge as a result of being offered access to the mNutrition content.

The wealth interactions are all positively signed—indicating that females and males from less wealthy households had higher IYCF knowledge benefits from the mNutrition service—but the point estimates are never statistically significantly different from zero and all three are less than 3% of the magnitude of the main effects. Thus, there is at most weak evidence of heterogeneity by wealth in the IYCF knowledge treatment effects.

**Table 8.8: Heterogeneity in mNutrition impacts on IYCF knowledge by household wealth**

| | Treatment | Likelihood of <150% NPL | Treatment x likelihood of <150% NPL | N |
|---|---|---|---|---|
| Percent of answers correct (female) | 0.850 | -0.089*** | 0.024 | 2,469 |
| | (0.619) | (0.016) | (0.022) | |
| Percent of answers correct (male) | 1.653** | -0.067*** | 0.008 | 2,204 |
| | (0.723) | (0.017) | (0.025) | |
| Anderson Index: combined percent correct of females and males | 0.095** | -0.005*** | 0.001 | 2,114 |
| | (0.040) | (0.001) | (0.001) | |

Notes: Estimates from the mNutrition Tanzania endline survey sample. Standard errors are in parentheses and clustered at the village level. Reported are impact estimates from the regressions of each outcome on the treatment indicator, likelihood of being below 150% of the NPL, and an interaction term of the latter two. The poverty likelihood is converted from the PPI using the threshold values created for Tanzania. The model controls for the value of the respective outcome at baseline. * p<0.10 **p<0.05 ***p<0.01

Source: Authors' own

Across all the primary and secondary outcomes there is no evidence that more wealthy households benefitted more from the service. We also find strong evidence that individuals in less wealthy households have lower IYCF knowledge, which would support the hypothesis that the mNutrition content could be more likely to duplicate existing knowledge for members of wealthier households, resulting in smaller impacts for these individuals. However, while the wealth interactions for the IYCF knowledge outcomes are of the correct sign as regards supporting this theory—indicating that the mNutrition service may have had larger impacts on IYCF knowledge for individuals from less wealthy households—the point estimates are small in magnitude and not statistically significantly different from zero. Given these weak results and a plethora of alternative mechanisms that could lead to the same wealth gradient in the treatment effects, we are unable to definitively state why the service had larger impacts for less wealthy households. Nevertheless, at a minimum, the findings do suggest that mobile message-based BCC may not be an unreasonable way to reach and affect the nutrition-related behaviours of vulnerable households in developing countries. This is important to highlight as there are a number of reasons why we might have expected to find larger impacts among wealthier households. For example, wealthier households could have been less constrained in regard to acquiring the food or other inputs suggested by the mNutrition messaging, household member literacy or imperfect message comprehension could have imposed more binding barriers on less wealthy households, and/or increased mobile phone attachment and use among the wealthy could have resulted in more exposure to the message content.

# 9 Impact on mobile phone use, information sources, trust, and household expenditure

Beyond the primary and secondary outcomes discussed in detail in previous sections, the detail of the mNutrition endline survey enables us to explore additional outcomes that the service may have affected. We focus on outcomes that are likely to be of interest to policymakers and MNOs, as well as those that may help us to better understand the mechanisms through which the service could have impacted the primary and secondary outcomes. Specifically, we test whether the offer of access to the mNutrition content changed how individuals use their mobile phones, whether the service affected respondents' sources of nutrition information and their trust in nutrition information received through different platforms or people, and whether the service had any impact on household expenditure overall, on food items, or on non-food items.

## 9.1 mNutrition and mobile phone use

The mobile phone use data were collected through separate interviews with the primary female and the primary male. Respondents were asked general questions about mobile phone access, ownership, use of mobile money or internet, and network participation, and more detailed questions about how they interacted with mobile phones in the last 14 days. From a household expenditure module included in the endline survey we also observed a measure of total household spending on mobile phone air time during the month preceding the survey. Table 9.1 presents the mNutrition impact estimates on these outcomes for primary females and Table 9.2 does the same for primary males. All impact estimates, when possible, are based on ANCOVA versions of the basic model, and otherwise are based on simple differences versions of the basic model.

### 9.1.1 Primary female phone use

The estimates in Table 9.1 suggest there is no impact of the offer of access to mNutrition on the likelihood that women own a mobile phone, on whether they need to travel more than 30 minutes to the nearest place they can charge their mobile phone, on the amount of money they spend to charge their mobile phone in a typical month, or on the network provider of the main mobile phone.

However, it is clear from Table 9.1 that the mNutrition offer had important impacts on how women interact with their mobile phones. They reported being 5.3 percentage points more likely to have used their mobile phone to make calls (p-value 0.016), 4.0 percentage points more likely to have received calls (p-value 0.074), 6.2 percentage points more likely to have written text messages (p-value <0.01), 6.3 percentage points more likely to have received text messages (p-value <0.01), 4.5 percentage points more likely to have received mobile money (p-value 0.01), 3.3 percentage points less likely to not have used their mobile phone during the past 14 days (p-value 0.07), and 6.1 percentage points more likely to have used their mobile on most days during the past 14 days (0.02). These effects all suggest that access to the mNutrition service increased the frequency with which females interact with their mobile phones, often leading to large increases relative to the control group means. The impact on the likelihood of having used a mobile in the past 14 days to write (read) a text message is equivalent to 13.6 (11.2%) of the control group mean. Furthermore, the implied effect on the likelihood of receiving mobile money is even larger, suggesting mNutrition led to an increase equivalent to 26.5% of the control group mean.

Although the point estimate for the impact of the offer of access to mNutrition on the likelihood of having a mobile money account is positive, it is not statistically different from zero (p-value 0.132).

We similarly find no impact on the number of mobile phones owned or accessed by the respondent, and there is little variation in this outcome among the women in the sample: 2.67% of the women with a non-missing answer own or have access to no mobile phones, 88.25% own or have access to exactly one mobile phone, 9.04% own or have access to two mobile phones, and just one woman owns or has access to three mobile phones.

In addition to the null results mentioned above, we highlight here that we find no impacts of access to the mNutrition service on the length of time that respondents reported having had their main SIM card or a second SIM card. We also find no statistically significant impact on the likelihood that females reported having the same primary mobile phone number as they had at baseline or the same MNO for their main phone at endline as they did at baseline. These are important findings to emphasise as they do not support the theory that access to the mNutrition service affected individuals' retention of mobile phones or loyalty to an MNO. Critically, only 30.4% of females reported having the same main mobile phone number at baseline and at endline in the control group. This underlines one of the potential barriers to consistently reaching individuals through mobile phones in the study areas.

**Table 9.1: Impact estimates of mNutrition on primary female phone use, ANCOVA**

| | Control mean | Impact estimates, basic controls | Impact estimates, extended controls | N |
|---|---|---|---|---|
| Owns a mobile phone | 0.572 | 0.024 | 0.023 | 2,469 |
| | | (0.021) | (0.020) | |
| Used mobile phone in the last 14 days to make calls | 0.625 | 0.053** | 0.048** | 2,469 |
| | | (0.022) | (0.020) | |
| Used mobile phone in the last 14 days to receive calls | 0.701 | 0.040* | 0.036* | 2,469 |
| | | (0.022) | (0.020) | |
| Used mobile phone in the last 14 days to write text messages | 0.457 | 0.062*** | 0.059*** | 2,469 |
| | | (0.021) | (0.019) | |
| Used mobile phone in the last 14 days to receive text messages | 0.564 | 0.063*** | 0.058*** | 2,469 |
| | | (0.021) | (0.020) | |
| Used mobile phone in the last 14 days to send mobile money | 0.092 | 0.010 | 0.008 | 2,469 |
| | | (0.014) | (0.012) | |
| Used mobile phone in the last 14 days to receive mobile money | 0.170 | 0.045** | 0.041*** | 2,469 |
| | | (0.017) | (0.016) | |
| Used mobile phone in the last 14 days to use mobile internet | 0.064 | -0.000 | -0.002 | 2,469 |
| | | (0.010) | (0.009) | |
| No mobile phone use in the last 14 days | 0.209 | -0.033* | -0.029* | 2,469 |
| | | (0.018) | (0.017) | |
| Mobile phone used on most days over the last 14 days | 0.448 | 0.061** | 0.053** | 2,064 |
| | | (0.025) | (0.021) | |
| Charges phone at home | 0.682 | 0.044 | 0.034 | 1,454 |
| | | (0.032) | (0.030) | |
| Takes less than 30 minutes to get to the nearest place to charge their phone | 0.983 | -0.009 | -0.008 | 374 |
| | | (0.015) | (0.016) | |
| Amount spent on charging phone in an average month (TSH) | 1,378.361 | -172.536 | -83.379 | 374 |
| | | (141.268) | (129.729) | |

| | | | | |
|---|---|---|---|---|
| Has used phone for mobile internet (e.g. WhatsApp, Facebook, email) | 0.114 | 0.003 | 0.002 | 2,469 |
| | | (0.015) | (0.014) | |
| Has mobile money account | 0.572 | 0.035 | 0.036 | 2,469 |
| | | (0.023) | (0.022) | |
| Number of phones owned or accessed by primary female | 1.060 | 0.008 | 0.007 | 2,468 |
| | | (0.015) | (0.015) | |
| Network provider: Airtel | 0.023 | -0.001 | 0.002 | 2,402 |
| | | (0.014) | (0.015) | |
| Network provider: Halotel | 0.103 | -0.016 | -0.000 | 2,402 |
| | | (0.032) | (0.030) | |
| Network provider: Vodacom | 0.718 | -0.005 | -0.027 | 2,402 |
| | | (0.047) | (0.041) | |
| Network provider: Smart | 0.156 | 0.021 | 0.024 | 2,402 |
| | | (0.037) | (0.035) | |
| Network provider: Tigo | 0.000 | 0.002 | 0.002 | 2,402 |
| | | (0.001) | (0.001) | |
| Length of ownership (in months), first SIM | 34.910 | -0.598 | -0.922 | 1,755 |
| | | (2.167) | (2.078) | |
| Length of ownership (in months), second SIM | 25.598 | -1.674 | -2.000 | 179 |
| | | (3.741) | (3.959) | |
| Doesn't know when acquired SIM card | 0.275 | -0.009 | -0.014 | 2,398 |
| | | (0.021) | (0.020) | |
| Main phone number the same at baseline and endline | 0.304 | 0.028 | 0.020 | 2,396 |
| | | (0.024) | (0.023) | |
| Main phone number uses the same MNO at baseline and endline | 0.749 | 0.020 | 0.009 | 2,396 |
| | | (0.027) | (0.025) | |

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 the treatment from an OLS regression of the outcome of interest on the treatment variable, controlling for baseline household classification and value of the respective outcome at baseline. For variables that were not collected at baseline, the model controls for phone ownership 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, and PPI score. Control mean is comparison group's mean at endline. Estimates for outcomes that were not observed at baseline are based on simple differences. * p<0.10 **p<0.05 ***p<0.01

Source: Authors' own

### 9.1.2 Primary male phone use

Table 9.2 continues by presenting the same set of results for primary males. As was the case for females, we find no impact of access to mNutrition on the likelihood that males own a mobile phone, on whether men need to travel more than 30 minutes to the nearest place where they can charge their mobile phone, on the amount of money men spend to charge their mobile phone in a typical month, or on the network provider for their main mobile phone.

In contrast with females, for whom we do find an impact on the likelihood of having used a mobile phone to make and receive calls in the last 14 days, we do not find a statistically significant impact of the randomly assigned offer of access to mNutrition on the likelihood that males used a mobile phone to make or receive calls. We do, however, still find large and statistically significant impacts of access to mNutrition on the self-reported likelihood that men used their mobile phone to write or receive text messages: men were 6.4 percentage points (9.6% of the control group mean) more likely to have used their mobile phone to write a text message and 6.9 percentage points (8.8% of the control group mean) more likely to have used their mobile phone to receive a text message during the previous 14 days.

Interestingly, although there is no statistically significant impact of access to mNutrition on the likelihood that men reported using their mobile phone to receive mobile money, we do estimate a statistically significant impact on the likelihood of men using their mobile phone to send mobile money during the previous 14 days. The point estimate indicates that the offer of access to mNutrition increased the likelihood that men reported using their mobile phone to send mobile money by 5.4 percentage points, or 19.0% of the control group mean. While we did not collect information on who respondents were receiving mobile money from or sending mobile money to, we find impacts of mNutrition on the likelihood that females received mobile money and impacts on the likelihood that males sent mobile money, both of similar magnitudes. At a minimum, this hints that the service's impacts on mobile money use may be affected by an increased likelihood of mobile money transfers from the primary males to primary females in the treatment group, though we do not have adequate data to show that this type of intra-household mobile money transfer is occurring.

As was the case for the primary females, we find no statistically significant impacts of access to the mNutrition service on the length of time males have had their main SIM or a second SIM, on the likelihood that the primary males had the same main mobile phone number at baseline and at endline, or on the likelihood that the main phone number of the primary male used the same network at baseline and at endline. Less than half of the primary males (47.5%) had the same mobile number at baseline and at endline, though 80.0% of primary males reported that their main mobile phone used the same network at baseline and at endline. These non-results provide further support for the idea that, while the mNutrition service increased the frequency of mobile phone use in treatment households, it did not increase loyalty to a particular mobile network or increase individuals' attachment to their mobile phones.

**Table 9.2: Impact estimates of mNutrition on primary male phone use, ANCOVA**

| | Control mean | Impact estimates, basic controls | Impact estimates, extended controls | N |
|---|---|---|---|---|
| Owns a mobile phone | 0.892 | 0.006 | 0.001 | 1,210 |
| | | (0.019) | (0.018) | |
| Used mobile phone in the last 14 days to make calls | 0.856 | 0.028 | 0.023 | 1,210 |
| | | (0.021) | (0.020) | |
| Used mobile phone in the last 14 days to receive calls | 0.905 | 0.010 | 0.007 | 1,210 |
| | | (0.017) | (0.016) | |
| Used mobile phone in the last 14 days to write text messages | 0.665 | 0.064** | 0.056** | 1,210 |
| | | (0.025) | (0.024) | |
| Used mobile phone in the last 14 days to receive text messages | 0.787 | 0.069*** | 0.059*** | 1,210 |
| | | (0.023) | (0.022) | |
| Used mobile phone in the last 14 days to send mobile money | 0.285 | 0.054** | 0.046* | 1,210 |
| | | (0.027) | (0.026) | |
| | 0.338 | 0.027 | 0.017 | 1,210 |

| | | | | |
|---|---|---|---|---|
| Used mobile phone in the last 14 days to receive mobile money | | (0.030) | (0.029) | |
| Used mobile phone in the last 14 days to use mobile internet | 0.111 | -0.003 | -0.002 | 1,210 |
| | | (0.019) | (0.019) | |
| No mobile phone use in the last 14 days | 0.059 | -0.020 | -0.016 | 1,210 |
| | | (0.013) | (0.013) | |
| Mobile phone used on most days over the last 14 days | 0.762 | 0.041 | 0.037 | 1,171 |
| | | (0.028) | (0.026) | |
| Charges phone at home | 0.700 | 0.049 | 0.040 | 1,079 |
| | | (0.031) | (0.030) | |
| Takes less than 30 minutes to get to the nearest place to charge their phone | 0.987 | -0.017 | -0.019 | 263 |
| | | (0.015) | (0.015) | |
| Amount spent on charging phone in an average month (TSH) | 1,295.098 | 289.023 | 278.727 | 263 |
| | | (183.269) | (169.434) | |
| Has used phone for mobile internet (e.g. WhatsApp, Facebook, email) | 0.213 | 0.001 | 0.001 | 1,210 |
| | | (0.025) | (0.024) | |
| Has mobile money account | 0.850 | 0.013 | 0.006 | 1,210 |
| | | (0.023) | (0.022) | |
| Number of phones owned or accessed by primary female | 1.154 | 0.011 | 0.014 | 1,210 |
| | | (0.026) | (0.026) | |
| Network provider: Airtel | 0.026 | 0.003 | 0.005 | 1,196 |
| | | (0.013) | (0.013) | |
| Network provider: Halotel | 0.110 | -0.020 | -0.006 | 1,196 |
| | | (0.035) | (0.032) | |
| Network provider: Vodacom | 0.727 | -0.011 | -0.032 | 1,196 |
| | | (0.053) | (0.048) | |
| Network provider: Smart | 0.000 | 0.002 | 0.002 | 1,196 |
| | | (0.002) | (0.002) | |
| Network provider: Tigo | 0.134 | 0.028 | 0.032 | 1,196 |
| | | (0.044) | (0.042) | |
| Length of ownership (in months), 1st SIM | 55.621 | -1.792 | -2.774 | 1,009 |
| | | (3.754) | (3.748) | |
| Length of ownership (in months), 2nd SIM | 44.877 | -5.946 | -2.869 | 176 |
| | | (6.321) | (6.654) | |
| Doesn't know when acquired SIM card | 0.155 | -0.007 | -0.008 | 1,196 |
| | | (0.026) | (0.025) | |
| Main phone number the same at baseline and endline | 0.475 | 0.039 | 0.024 | 1,194 |
| | | (0.037) | (0.037) | |
| Main phone number uses the same MNO at baseline and endline | 0.800 | -0.010 | -0.022 | 1,194 |
| | | (0.034) | (0.032) | |

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 the treatment from an OLS regression of the outcome of interest on the treatment variable, controlling for baseline household classification and value of the respective outcome at baseline. For variables that were not collected at baseline, the model controls for phone ownership 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, and PPI score. Control mean is comparison group's mean at endline. Estimates for outcomes that were not observed at baseline are based on simple differences. * $p<0.10$ ** $p<0.05$ *** $p<0.01$

Source: Authors' own

## 9.2 Household expenditure

In this subsection we present mNutrition estimates of the impact of mNutrition access on the log of per capita total annual household expenditure and on total household expenditure on mobile phone vouchers during the previous month. The expenditure information was collected through two survey modules, one of which was focused on non-food expenditure and a second focused on food consumption and expenditure in the week preceding the survey. The non-food expenditure module was asked of the primary male or primary female in all sample households, while the food consumption and expenditure module was asked of the primary female or the household member responsible for preparing food in the household.

Due to budget and time constraints, we were unable to field a full-length expenditure module. Instead, we follow guidance provided in Beegle *et al*. (2012) and used a shortened expenditure module that asked only about the 16 non-food items that were the most important expenditure categories for households in rural Iringa in the most recent (2011–2012) Tanzania Household Budget Survey (THBS) (Tanzania National Bureau of Statistics, 2012). We additionally asked about expenditures on clothing and footwear for men, women, and children. On average, these 16 items plus expenditure on clothing and footwear made up 84.6% of households' total expenditure on non-food items. Enumerators asked households about their total spending on each of these items in the last month (for items typically purchased more frequently) or in the last year (for larger, less frequent purchases).

For the food items, we similarly asked about household consumption and expenditure for the 16 food items that made up the largest share of the household food budget for rural households in Iringa according to the same data. Collectively, these 16 items constituted 85.6% of the total household food budget. For each food item, the respondent was asked how much the household consumed during the seven days preceding the survey. Conditional on the household having consumed a positive amount, they were then asked how much of that total was from own production and how much was purchased. If any of the items was purchased, the household was asked how much was purchased and how much they paid for that amount. We use the prices that households report paying for each food item to value the full consumption of that item.[22] This assumes that the value of each consumed item is the same, regardless of whether the item was purchased or self-produced. This may be a strong assumption, but it should not differentially affect the expenditure for treatment and control households. It would have to be true that the mNutrition content induced households to change agricultural production or the likelihood that households consumed items out of own production as opposed to out of own production. Given the content of the messages—which had little to do with agriculture or the relative benefits of self-produced compared to market goods—this seems unlikely in the context of the study.

For the non-food and food items we asked directly about during the expenditure module, the responses are used directly as measures of weekly, monthly, or annual expenditure on that item.

[22] In cases where the household consumed an item but did not report having purchased any of that item in the previous week, we used the median price paid in the village to value the household's consumption of that item. If no households in the village purchased that item, then we used the median price paid in the household's ward to value the consumption of that item. If no households in the ward purchased that item, then we used the median price paid in the household's district to value the consumption of that item, and if no households in the district purchased that item then we used the median price paid in the full sample to value the consumption of that item.

However, to get a measure of total household expenditure, we inflate the total reported expenditure amount using the inverse of the average budget share accounted for by the items included in the expenditure module, with the budget share information taken from the THBS 2011/12. To allow for the possibility that the service may have affected spending on mobile phone vouchers, clothing, or footwear, we exclude the budget shares for these items from the total non-food spending, inflate the remaining observed total expenditure, and then add back in spending on these items.[23]

Table 9.3 presents the impact estimates using the log of total per capita expenditure (inflated as described above) as an outcome, as well as for the total monthly reported household spending on mobile phone vouchers. We show simple differences estimates from both the basic model as well as from the extended model, as described in detail in Section 6, for log total household expenditure per capita and ANCOVA estimates for total monthly expenditure on mobile phone vouchers as we collected information on mobile phone spending at baseline. On average, households in the control group have a log per capita annual consumption of 13.6 and a total annual household expenditure per capita of TSH 752,794.3. There is no impact of access to the mNutrition service on log per capita annual household spending in either model, and t-statistics are below 1 in both columns.

Mobile phone vouchers are an important expense for the sample households, with control households spending, on average, TSH 5,295.5 on them per month. This represents roughly 1.5% of the total household annual budget. Consistent with the evidence presented in Table 9.1 and Table 9.2, the random offer of access to the service statistically significantly increased household spending on mobile phone vouchers: by TSH 510.5, on average. This represents an increase in mobile phone voucher spending equivalent to 9.6% of the control group mean and is statistically significant at the 10% level (p-value 0.079). This impact is consistent with the primary female and primary male self-reports, which suggested that mobile phone use increased as a result of access to the service. Converted to annual spending, the impact estimates indicate that treated households spend, on average, TSH 6,126.4 more on mobile phone use.

Although we are not able to allocate mobile phone spending to mobile networks, and therefore are not able to definitively identify what fraction of the increased mobile phone activity and spending occurred on the mobile network of the phone registered for the mNutrition service (for treatment households), we can use the data collected to come up with a back-of-the-envelope estimate for this quantity. To do so, we use the fact that households typically have just one mobile phone number—80.7% of households own at most one phone number—and this figure does not vary with treatment status. Further, 74.9% of primary females and 80.0% of primary males had the same MNO for their main mobile phone at baseline and at endline. We, conservatively, assume that mobile phone spending is spread equally across mobile phone numbers, that each number belongs to a different network, and that if a female had a main mobile phone on a different network at baseline and at endline then that household does not retain a mobile phone on the baseline main network. Under these assumptions, the incremental spending on the network of the main mobile phone can be estimated as $(1 - 0.749) \times 0 + 0.749 \times 0.807 \times 510.5 = 308.6\, TSH$ per month, where $510.5$ is the treatment effect of mNutrition on household mobile phone spending per month and $0.749 \times 0.807 = 0.604$ is the fraction of that effect that we can allocate to the main mobile network at baseline. Converting this to an annual amount implies that we expect the service

[23] Excluding clothing, footwear, and mobile phone vouchers, the other items included in the non-food expenditure module account for 81.8% of total non-food and non-clothing, footwear, or mobile phone voucher spending for households in rural Iringa in the 2011/12 THBS. We therefore multiply the total non-food expenditure (excluding these categories) by 1/.818 and then add back in the observed spending on these categories to get an estimate of total non-food expenditure. Similarly, total reported food expenditure/consumption is inflated by multiplying by 1/.855 to arrive at the total food expenditure/consumption measure.

to increase household spending on mobile phone use for the network of the main mobile number by TSH 3,703.1 per year.

**Table 9.3: Impact estimates of mNutrition on household expenditures**

| | Control mean | Impact estimates, basic controls | Impact estimates, extended controls | N |
|---|---|---|---|---|
| Log per capita: total annual value of household consumption (TSH), inflated | 13.353 | 0.019 | 0.023 | 2,589 |
| | | (0.035) | (0.028) | |
| Household expenditure on phone vouchers last month (TSH) | 5,295.503 | 510.536* | 498.293* | 2,595 |
| | | (289.045) | (263.273) | |

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 the treatment from an OLS regression of the outcome of interest on the treatment variable, controlling for baseline household classification. 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, and PPI score. Annual values of consumption are constructed from households' reported consumption of food over the last seven days and non-food over last 30 days or 12 months depending on item. Where prices for own consumption are not known, they are imputed from prices reported on the lowest administrative level available. Reported total spending is inflated using budget shares from the 2011/12 THBS in order to predict total household expenditure given the restricted set of the 32 most important budget items included in the endline survey. Control mean is comparison group's mean at endline. * p<0.10 **p<0.05 ***p<0.01

Source:  Authors' own

## 9.3 Information sources and trust in information

We continue by exploring whether the offer of access to the mNutrition service affected where individuals get their nutrition information from, what they identify as their main (up to two) sources of nutrition information, and whether they agree or strongly agree that they would feel confident in trusting 'breastfeeding, complementary feeding, or general health information' from different sources. As with the mobile phone use data discussed in Section 9.1, the data come from questions that were asked separately of the primary female and the primary male in each household. All impact estimates are based on the basic and extended ITT ANCOVA models.

### 9.3.1 Primary female

Table 9.4 shows the impact estimates of the offer of access to the mNutrition service on the information sources, main information sources, and trust in information from different sources outcomes for primary females. Primary females in treatment households are 23.0 percentage points more likely to report automatic text messages as a source of nutrition information than primary females in control households (p-value <0.01). This represents an 80.4% increase relative to the control group mean. Females in treatment households are also statistically significantly more likely to report their spouse, TV/radio/posters, and community health workers as sources of nutrition information. However, relative to the impact on automatic text messages being a source of information, the impacts for the other sources are substantially smaller in magnitude, both in absolute terms and relative to the respective control group means (which are two to three times larger for spouse, TV/radio/posters, and community health workers than they are for automatic text messages).

Primary females are 10.4 percentage points more likely to report that automatic text messages are one of two primary sources for nutrition information (p-value < 0.01), an effect size that is more than three times the control group mean. This implies that primary females in treatment households are 4.15 times more likely to identify automatic text messages as their control group counterparts, on average. Government health workers remain the most commonly mentioned primary source of nutrition information, with 97.8% of treatment and control households identifying them in their response. The only other statistically significant effect for the primary source of nutrition information questions is for TV/radio/posters, which suggests that access to the service decreases the likelihood that TV/radio/posters are identified as a primary source by 5.7 percentage points (roughly 20% of the control group mean). This potentially indicates that the service induced women to substitute away from information received on TV/radio/posters and toward the information contained in the mNutrition content.

Females in treatment households are also 4.1 percentage points more likely to indicate that they would be confident in the breastfeeding, complementary feeding, or general health information they received through automatic text messages. However, this point estimate is quite small relative to the control group mean of 83.2%, which reinforces a finding from the baseline data collection that households—prior to exposure to the service—were already open and likely to trust the nutrition information they received through text messages. Treatment females additionally reported that they would be confident in information on these topics they received from friends or neighbours, government health workers, private health clinics or hospitals, or community health workers.

**Table 9.4: Impact estimates of mNutrition on female information sources and trust, ANCOVA**

| | Control mean | Impact estimates, basic controls | Impact estimates, extended controls | N |
|---|---|---|---|---|
| Spouse is a source of information | 0.619 | 0.044* | 0.038* | 1,889 |
| | | (0.024) | (0.022) | |
| Family is a source of information | 0.486 | 0.029 | 0.024 | 2,469 |
| | | (0.023) | (0.021) | |
| Friends/neighbours are a source of information | 0.550 | 0.016 | 0.011 | 2,469 |
| | | (0.024) | (0.023) | |
| Automated text messages are a source of information | 0.286 | 0.230*** | 0.209*** | 2,468 |
| | | (0.027) | (0.021) | |
| Government health workers are a source of information | 0.978 | -0.000 | -0.000 | 2,469 |
| | | (0.007) | (0.007) | |
| Non-government health facilities are a source of information | 0.568 | 0.026 | 0.006 | 2,469 |
| | | (0.032) | (0.026) | |
| TV/radio/posters are a source of information | 0.771 | 0.046** | 0.044** | 2,469 |
| | | (0.020) | (0.019) | |
| Traditional health workers are a source of information | 0.050 | 0.005 | 0.005 | 2,469 |
| | | (0.009) | (0.009) | |
| Non-government health workers are a source of information | 0.434 | 0.045 | 0.032 | 2,469 |
| | | (0.028) | (0.026) | |
| Community health workers are a source of information | 0.794 | 0.046** | 0.041** | 2,469 |
| | | (0.021) | (0.020) | |

| | | | | |
|---|---|---|---|---|
| Spouse is the primary source of information | 0.310 | -0.004 | -0.006 | 2,469 |
| | | (0.026) | (0.025) | |
| Family is the primary source of information | 0.125 | -0.019 | -0.016 | 2,469 |
| | | (0.018) | (0.017) | |
| Friends/neighbours are the primary source of information | 0.090 | -0.017 | -0.015 | 2,469 |
| | | (0.017) | (0.016) | |
| Automated text messages are the primary source of information | 0.033 | 0.104*** | 0.100*** | 2,469 |
| | | (0.017) | (0.016) | |
| Government health workers are the primary source of information | 0.990 | 0.033 | 0.025 | 2,469 |
| | | (0.030) | (0.028) | |
| Non-government health facilities are the primary source of information | 0.277 | 0.023 | 0.014 | 2,469 |
| | | (0.036) | (0.035) | |
| TV/radio/posters are the primary source of information | 0.287 | -0.057* | -0.057** | 2,469 |
| | | (0.030) | (0.029) | |
| Traditional health workers are the primary source of information | 0.007 | -0.005 | -0.004 | 2,469 |
| | | (0.004) | (0.004) | |
| Non-government health workers are the primary source of information | 0.038 | -0.012 | -0.012 | 2,469 |
| | | (0.009) | (0.009) | |
| Community health workers are the primary source of information | 0.721 | -0.024 | -0.011 | 2,469 |
| | | (0.051) | (0.049) | |
| Agree they would trust information from their spouse | 0.894 | 0.013 | 0.008 | 1,890 |
| | | (0.014) | (0.014) | |
| Agree they would trust information from their family | 0.763 | 0.014 | 0.010 | 2,469 |
| | | (0.021) | (0.020) | |
| Agree they would trust information from their friends/neighbours | 0.642 | 0.047* | 0.043* | 2,469 |
| | | (0.025) | (0.023) | |
| Agree they would trust information from automated text messages | 0.832 | 0.041** | 0.042** | 2,469 |
| | | (0.017) | (0.017) | |
| Agree they would trust information from a government health worker | 0.984 | 0.008* | 0.008* | 2,469 |
| | | (0.005) | (0.004) | |
| Agree they would trust information from a private clinic/hospital | 0.919 | 0.022* | 0.023** | 2,469 |
| | | (0.012) | (0.011) | |
| Agree they would trust information from TV/radio/posters | 0.890 | 0.014 | 0.016 | 2,469 |
| | | (0.014) | (0.014) | |
| Agree they would trust information from a traditional health worker | 0.091 | -0.005 | -0.002 | 2,469 |
| | | (0.012) | (0.012) | |
| Agree they would trust information from a non-government health worker | 0.841 | 0.015 | 0.013 | 2,469 |
| | | (0.016) | (0.015) | |
| Agree they would trust information from a community health worker | 0.961 | 0.013* | 0.012* | 2,469 |
| | | (0.007) | (0.007) | |

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 the treatment from an OLS regression of the outcome of interest on the 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, and PPI score. Control mean is comparison group's mean at endline. * p<0.10 **p<0.05 ***p<0.01

Source: Authors' own

### 9.3.2 Primary male

Table 9.5 continues by presenting the same set of impact estimates for the primary males in the sample. In general, males seem less likely to change their sources of nutrition information than females. Still, we estimate that the offer of access to the mNutrition service increased the likelihood that automatic text messages are a source of nutrition information by 17.9 percentage points (54.1% of the control group mean) and males are 10.0 percentage points more likely to identify automatic text messages as a primary source of nutrition information (192.3% of the control group mean). Treatment males are also marginally statistically significantly less likely to identify government health workers and TV/radio/posters as primary sources of nutrition information, which may suggest that mNutrition crowded out these alternative information sources for males. We find no impact on the likelihood that males would trust breastfeeding, complementary feeding, or general health information from any of the sources.

**Table 9.5: Impact estimates of mNutrition on male information sources and trust, ANCOVA**

| | Control mean | Impact estimates, basic controls | Impact estimates, extended controls | N |
|---|---|---|---|---|
| Spouse is a source of information | 0.811 | 0.024 | 0.016 | 1,829 |
| | | (0.019) | (0.017) | |
| Family is a source of information | 0.479 | 0.025 | 0.009 | 2,204 |
| | | (0.026) | (0.023) | |
| Friends/neighbours are a source of information | 0.575 | -0.021 | -0.030 | 2,204 |
| | | (0.025) | (0.024) | |
| Automated text messages are a source of information | 0.331 | 0.179*** | 0.158*** | 2,204 |
| | | (0.029) | (0.024) | |
| Government health workers are a source of information | 0.889 | -0.008 | -0.008 | 2,204 |
| | | (0.015) | (0.015) | |
| Non-government health facilities are a source of information | 0.537 | 0.019 | 0.003 | 2,204 |
| | | (0.029) | (0.026) | |
| TV/radio/posters are a source of information | 0.877 | -0.013 | -0.019 | 2,204 |
| | | (0.016) | (0.015) | |
| Traditional health workers are a source of information | 0.057 | 0.008 | 0.007 | 2,203 |
| | | (0.012) | (0.013) | |
| Non-government health workers are a source of information | 0.443 | 0.037 | 0.022 | 2,204 |
| | | (0.027) | (0.024) | |
| Community health workers are a source of information | 0.741 | 0.027 | 0.015 | 2,204 |
| | | (0.024) | (0.023) | |
| Spouse is the primary source of information | 0.606 | -0.023 | -0.021 | 2,203 |
| | | (0.035) | (0.034) | |
| Family is the primary source of information | 0.091 | -0.027 | -0.029 | 2,203 |
| | | (0.018) | (0.018) | |
| Friends/neighbours are the primary source of information | 0.066 | -0.005 | -0.004 | 2,203 |
| | | (0.015) | (0.015) | |

| | | | | |
|---|---|---|---|---|
| Automated text messages are the primary source of information | 0.052 | 0.100*** | 0.099*** | 2,203 |
| | | (0.020) | (0.020) | |
| Government health workers are the primary source of information | 0.897 | -0.062* | -0.064* | 2,203 |
| | | (0.036) | (0.036) | |
| Non-government health facilities are the primary source of information | 0.211 | 0.033 | 0.029 | 2,203 |
| | | (0.032) | (0.031) | |
| TV/radio/posters are the primary source of information | 0.452 | -0.057* | -0.050 | 2,203 |
| | | (0.033) | (0.033) | |
| Traditional health workers are the primary source of information | 0.008 | -0.004 | -0.004 | 2,203 |
| | | (0.004) | (0.005) | |
| Non-government health workers are the primary source of information | 0.036 | 0.003 | 0.003 | 2,203 |
| | | (0.010) | (0.011) | |
| Community health workers are the primary source of information | 0.430 | 0.057 | 0.053 | 2,203 |
| | | (0.039) | (0.038) | |
| Agree they would trust information from their spouse | 0.945 | 0.011 | 0.007 | 1,830 |
| | | (0.013) | (0.012) | |
| Agree they would trust information from their family | 0.785 | 0.001 | -0.008 | 2,204 |
| | | (0.022) | (0.021) | |
| Agree they would trust information from their friends/neighbours | 0.658 | 0.021 | 0.010 | 2,204 |
| | | (0.029) | (0.026) | |
| Agree they would trust information from automated text messages | 0.853 | 0.021 | 0.020 | 2,204 |
| | | (0.019) | (0.019) | |
| Agree they would trust information from a government health worker | 0.987 | 0.002 | 0.001 | 2,204 |
| | | (0.004) | (0.004) | |
| Agree they would trust information from a private clinic/hospital | 0.932 | 0.013 | 0.012 | 2,204 |
| | | (0.013) | (0.012) | |
| Agree they would trust information from the TV/radio/posters | 0.916 | -0.003 | -0.003 | 2,204 |
| | | (0.014) | (0.014) | |
| Agree they would trust information from a traditional health worker | 0.083 | 0.010 | 0.011 | 2,204 |
| | | (0.014) | (0.014) | |
| Agree they would trust information from a non-government health worker | 0.892 | 0.003 | 0.002 | 2,204 |
| | | (0.015) | (0.015) | |
| Agree they would trust information from a community health worker | 0.973 | -0.008 | -0.012 | 2,204 |
| | | (0.009) | (0.008) | |

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 the treatment from an OLS regression of the outcome of interest on the treatment variable, controlling for baseline household classification and value of the respective outcome at baseline for primary female, since variables on male's trust were not collected at baseline. Extended controls are covariates from baseline: household size, whether household head is female, age of individual responding, whether individual responding is literate in Swahili, PPI score, and whether individual owns a mobile phone. Control mean is comparison group's mean at endline. * p<0.10 **p<0.05 ***p<0.01

Source: Authors' own

## 9.4 Summary

Section 9 has highlighted the impacts of the offer of access to the mNutrition service on other outcomes related to mobile phone use and expenditure, general and primary sources of nutrition information, and trust in nutrition information received through different mediums. For both females and males, the service appears to have increased mobile phone use. We find meaningful impacts on the likelihood of sending and receiving text messages (for both males and females), and of making or receiving phone calls (for females), and we find an interesting and asymmetric pattern of results indicating an increased likelihood of receiving mobile money (for females) and an increased likelihood of sending mobile money (for males).

Entirely consistent with the increased use of mobile phones, treatment households also reported spending more on mobile phone vouchers. The impact estimates indicate that the offer of access to the service increased mobile phone spending by just under 10% during the previous month. While definitively allocating this increased mobile phone expenditure across mobile networks is not feasible given the data we collected, we show that even using conservative assumptions about how mobile phone vouchers are allocated by households indicates a meaningful increase in mobile phone spending on the main mobile network as a result of the mNutrition service.

In terms of how access to the mNutrition content affected where individuals get their nutrition-related information from, we find strong evidence that males and females in treatment households are much more likely to report getting nutrition information and their main nutrition information through automatic text messages. For individuals of both sexes, the service appears to have reduced the likelihood that respondents would identify TV/radio/posters as a source or a primary source of nutrition information.

In addition to impacting several of the primary and secondary study outcomes, the offer of access to the mNutrition service therefore also appears to have shaped individual and household behaviour in interesting ways.

# 10 Conclusions and Policy Implications

## 10.1 Conclusions

The rapid expansion of access to and the use of mobile phones across the developing world is an exciting development for policymakers. Societal issues associated with gaps in information—either low levels of knowledge or the high prevalence of inaccurate information—can potentially be addressed in a relatively inexpensive way by using mobile phones as a platform to disseminate knowledge. The possible benefits may be especially large for households located in remote and rural areas, who are more difficult to reach through more conventional information campaigns. Further, if effective, the lower per-beneficiary cost of spreading information through mobile messaging could enable governments and international aid organisations to target larger and more isolated populations.

However, while appealing in theory, the usefulness of mobile phones for disseminating important information, such as that related to health and nutrition behaviour, is untested in developing countries. Pessimism about the usefulness of mobile phones for the sharing of important information is well justified: low mobile phone attachment, poor network service, low levels of trust in non-human sources, and low message salience could all present insurmountable barriers to reaching households through mobile phone messaging. No matter how low the cost of developing and providing an intervention through mobile phones, if the intervention does not generate benefits for recipient individuals and households then other methods of BCC will be more efficient policy levers.

The purpose of the mNutrition impact evaluation is to address this ambiguity and fill the knowledge gap around the feasibility of using mobile phones to spread information. The evaluation therefore assesses the impact, cost-effectiveness, and commercial viability of the mNutrition service. The evaluation is being conducted by a consortium of researchers from Gamos, IDS, and IFPRI. The study team draws on a variety of quantitative and qualitative methods to gather and synthesise evidence about the mNutrition service as implemented in Tanzania.

This report presents the quantitative portion of the overall impact evaluation, relying on a cRCT to estimate the impact of the service on child anthropometry, IYCF practices, women's dietary diversity, and knowledge and beliefs about IYCF practices. In a sample of 180 randomly selected rural villages in Iringa region of Tanzania, the quantitative evaluation team randomly assigned eligible households in 90 villages to receive an offer of access to the mNutrition service. In the remaining 90 villages no similar offer was made.

In tandem with the cRCT evaluation design, the quantitative evaluation team relies on two rounds of household surveys, which were conducted by OPMT two years apart. In the baseline survey conducted between October and December 2016, OPMT interviewed 2,833 households across the 180 study villages. Just over 91% (2,595 households) were resurveyed during the endline survey between October and December 2018.

**The results 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. When asked about mNutrition service receipt specifically, this gap declines to 29.9 percentage points, but remains statistically significant. Both take-up gaps, while meaningful, are smaller than was expected during the study design phase. However, there is reason to believe that both versions of the take-up gap are underestimates of the true difference in exposure to the mNutrition service between treatment and control households. Administrative data on service registration for the main mobile phone numbers of treatment households from November 2017 indicated that a substantially higher share of treatment households was receiving the mNutrition messaging than the self-reported message receipt measures from the endline survey suggested. Although, given our data, it is not possible to more accurately characterise the take-up gap, under conservative assumptions about the administrative data and participation in control households the gap should approach 70 percentage points.

**Even among households that self-identified as service participants in the endline survey, receipt of the mNutrition content was relatively infrequent.** Partially, this may be due to intended variation in message spacing, whereby households were to receive fewer messages as children aged, combined with self-reports that give greater weight to recent months relative to earlier parts of the study period. 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 never noticed 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 sufficiently satisfied with the message content to implement the suggestions they contain. Encouragingly, **85.8% of females who ever 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 is 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.

Proceeding along the causal chain, **we find that males in treatment villages have statistically significantly higher IYCF knowledge scores, and that a measure of household-level IYCF knowledge similarly indicates that the mNutrition service increased knowledge of IYCF practices.** Males in treatment villages answer 1.7% more questions correctly on average, and treatment households have an IYCF knowledge index score that is 0.09 standard deviations higher. The impact on female IYCF knowledge is positively signed, but smaller and not statistically significantly different from zero. The finding of impacts on male knowledge but not female knowledge could be driven by the lower baseline level of male knowledge or by differences in access to the mobile phone that was registered for the service.

**These differences in IYCF knowledge translate into statistically significant differences in several key IYCF practices outcomes.** We estimate that children aged 6–35 months in treatment households consume a more diverse diet: the CDDS is 0.107 higher, on average, in treatment villages and treatment children are 3.8 percentage points more likely to meet the MDD standard. Children aged 6–23 months are also 6.9 percentage points more likely to meet the MAD standard. These all represent important differences in the diversity of children's diets in treatment villages. Interestingly, the service also increased the diversity of primary female diets, resulting in a 4.0 percentage point increase in the likelihood that primary females meet the MDD-W standard. **We find no evidence of statistically significant impacts on the other measured IYCF practices, including early initiation of breastfeeding, exclusive breastfeeding up to six months of age, or the introduction of solid, semi-solid, or soft foods between six and eight months of age.**

Although we estimate that the service statistically significantly improved IYCF knowledge and altered key IYCF behaviours in treatment households, **we find no evidence that the service affected child anthropometry.** The impact estimates are statistically indistinguishable from zero and small in magnitude for HAZ, WHZ, stunting, and wasting. This indicates that lack of information may not be the primary constraint to proper child nutrition in the study context, or that the observed changes in knowledge and behaviours were not large enough to meaningfully impact child anthropometry.

While we estimate that, on average, households that were offered access to the mNutrition content had higher values for a number of outcomes, **we do not find similar effects when comparing households that were randomly selected to receive the mNutrition content on the mobile phone of the primary female and the primary male to control households, or to treatment households that only received the content on the mobile of the primary female.** The point estimates for the households that received the content on both mobile phones (the T-F+M group) are, in general, not distinguishable from zero and occasionally suggest that these households perform worse than those assigned to just receive the content on the primary female's mobile phone (the T-F group). This could suggest that the sample households are able to efficiently share information with one another and therefore do not benefit from the second set of messages. Alternatively, there could be more complex intra-household relationship issues that drive down the response to messages when they are sent to both the primary female and the primary male.

Though not conclusive, we find some weak evidence that the service had larger impacts on child diets for households that were more likely to be below 150% of the Tanzanian NPL at baseline. However, we do not find similar evidence of heterogeneity in the impacts on IYCF knowledge.

**In addition to the primary and secondary outcomes, men and women in treatment villages interact significantly more with their mobile phones and are more likely to identify automatic text messages as a source of nutrition information**. Women are more likely to call, receive calls, text, receive texts, and receive mobile money if they are in a village that was randomly assigned to the treatment group, while men are more likely to send and receive text messages and to send mobile money. Consistent with the self-reports on increased mobile phone use, we find that treatment households spend 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 MNOs from providing the mNutrition service.

## 10.2 Policy Implications

Understanding whether or how the mNutrition service could be adapted to improve HAZ and stunting among children and have larger impacts on dietary diversity and other IYCF practices in beneficiary households is an important avenue for future research. While the quantitative evaluation has shown that there are important knowledge gaps in rural Iringa, and that access to the mNutrition content diminished these knowledge gaps, it also suggests that the barriers to linear growth in children are more complex than simple information constraints.

Pairing the mobile messaging with other interventions meant to overcome other, potentially binding barriers to proper nutrition seems like a promising path forward. For example, mNutrition content could be combined with in-person BCC meetings that target practices like breastfeeding that may be difficult to change solely through the provision of basic information. This may be especially important in order for mobile phone-based information to benefit the most vulnerable households and individuals, who may be less likely to have consistent access to a mobile phone, strong network services, resources for mobile phone charging and airtime, and the time and human capital to put into practice new information that they find useful. These barriers are highlighted as being salient for study individuals across the evaluation components (Barnett et al., 2020). Alternatively, if household resource constraints are a critical determinant of child undernutrition (which existing research indicates is the case), the service could be linked to resource transfers of cash or food in order to help households maximise the potential nutrition benefits of the transfer.

# References

Aker, J. C., Ghosh, I., and Burrell, J. (2016) 'The promise (and pitfalls) of ICT for agriculture initiatives'. *Agricultural Economics* (United Kingdom).

Alderman, H., Hoddinott, J., and Kinsey, B. (2006) 'Long term consequences of early childhood malnutrition', *Oxford Economic Papers*.

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., 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*.

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*.

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*.

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

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 health in Tanzania: Initial exploratory qualitative study report'.

Barnett, I., Batchelor, S., Faith, B., Gilligan, D., Gordon, J., Hidrobo, M., Mitchell, B., Palloni, G., & Scott, N. (2020). *External evaluation of mobile phone technology-based nutrition and agriculture advisory services in Africa: Final Tanzania mixed methods evaluation report*.

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*.

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*.

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*.

De Onis, M., Dewey, K. G., Borghi, E., A.W., O., M., B., B., D., … Branca, F. (2013) 'The world health organization’s global target for reducing childhood stunting by 2025: Rationale and proposed actions', *Maternal and Child Nutrition*.

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.

DHS (2016) *[Datasets]*, ICF International, Calverton, Maryland.

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

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*.

Food and Agriculture Organization of the United Nations (2018) *FAOSTAT Statistics Databas*e, UNFAO, 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', *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'.

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

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*.

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*.

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', *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*.

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

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*.

Jiang, H., Li, M., Wen, L. M., Hu, Q., Yang, D., He, G., … 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*.

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) '12 Common Applications and a Visual Framework', *Global Health: Science and Practice*.

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*.

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

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.*

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', *Systematic Reviews*.

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*.

Pop-Eleches, C., Thirumurthy, H., Habyarimana, J.P., Zivin, J.G., Goldstein, M.P., de Walque, D., MacKeen, L., Haberer, J., Kimaiyo, S., Sidle, J., Ngare, D., and Bangsberg, D.R. (2011) 'Mobile phone technologies improve adherence to antiretroviral treatment in a resource-limited setting: a randomized controlled trial of text message reminders', *AIDS* 25(6), pp. 825–834.

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*.

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

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*.

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*.

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

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

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

# 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 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[24]**:**

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

[24] 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 IRB approval

## B.1 IFPRI IRB

Please note: attachment to B1 removed to meet GDPR regulations.

## B.2 COSTECH IRB

Please note: attachment B2 removed to meet GDPR regulations.

# Annex C Supplementary tables

## C.1 Programme exposure

### Table 10.1: Primary female and spouse receipt of nutrition messages by treatment status

| | N | All | Treatment (T) | Control (C) | Normalised difference between (T) and (C) | P-value |
|---|---|---|---|---|---|---|
| Has received automatic text messages with nutrition information in past two years | 2,469 | 0.317 | 0.485 | 0.151 | 0.769 | 0.000 |
| | | (0.465) | (0.500) | (0.358) | | |
| Doesn't know if receives automatic text messages | 2,469 | 0.006 | 0.006 | 0.006 | 0.001 | 0.982 |
| | | (0.075) | (0.075) | (0.075) | | |
| Receives nutrition messages from government hospital | 2,469 | 0.006 | 0.006 | 0.006 | -0.009 | 0.830 |
| | | (0.078) | (0.075) | (0.080) | | |
| Receives nutrition messages from NGO/international organisation | 2,469 | 0.027 | 0.045 | 0.010 | 0.217 | 0.000 |
| | | (0.163) | (0.207) | (0.098) | | |
| Receives nutrition messages from church/mosque | 2,469 | 0.000 | 0.000 | 0.000 | | |
| | | (0.000) | (0.000) | (0.000) | | |
| Receives nutrition messages from Wazazi Nipendeni/15501 | 2,469 | 0.195 | 0.316 | 0.075 | 0.639 | 0.000 |
| | | (0.396) | (0.465) | (0.263) | | |
| Receives nutrition messages from 15014 number | 2,469 | 0.026 | 0.039 | 0.014 | 0.159 | 0.001 |
| | | (0.160) | (0.194) | (0.116) | | |
| Receives nutrition messages from an unknown source | 2,469 | 0.089 | 0.123 | 0.055 | 0.242 | 0.000 |
| | | (0.284) | (0.329) | (0.228) | | |
| Spouse has received automatic text messages with nutrition information in past two years | 1,998 | 0.160 | 0.268 | 0.055 | 0.603 | 0.000 |
| | | (0.366) | (0.443) | (0.228) | | |
| Doesn't know if spouse receives automatic text messages with nutrition information | 1,998 | 0.415 | 0.393 | 0.436 | -0.089 | 0.114 |
| | | (0.493) | (0.489) | (0.496) | | |
| Spouse receives nutrition messages from government hospital | 1,998 | 0.001 | 0.001 | 0.000 | 0.045 | 0.316 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | (0.022) | (0.032) | (0.000) | | |
| Spouse receives nutrition messages from NGO/international organisation | 1,998 | 0.011 | 0.017 | 0.005 | 0.118 | 0.022 |
| | | (0.104) | (0.130) | (0.070) | | |
| Spouse receives nutrition messages from Church/mosque | 1,998 | 0.000 | 0.000 | 0.000 | | |
| | | (0.000) | (0.000) | (0.000) | | |
| Spouse receives nutrition messages from Wazazi Nipendeni/15501 | 1,998 | 0.095 | 0.165 | 0.027 | 0.483 | 0.000 |
| | | (0.293) | (0.371) | (0.161) | | |
| Spouse receives nutrition messages from 15014 number | 1,998 | 0.011 | 0.016 | 0.005 | 0.111 | 0.028 |
| | | (0.102) | (0.127) | (0.070) | | |
| Spouse receives nutrition messages from unknown source | 1,998 | 0.057 | 0.095 | 0.021 | 0.321 | 0.000 |
| | | (0.232) | (0.293) | (0.142) | | |
| Has seen messages that spouse received from Wazazi Nipendeni | 143 | 0.874 | 0.875 | 0.870 | 0.016 | 0.944 |
| | | (0.333) | (0.332) | (0.344) | | |

Notes: Estimates from the mNutrition Tanzania endline survey sample. Standard deviations are in parentheses. P-values are from the tests of difference of means between treatment and control group.

Source: Authors' own

### Table 10.2: Primary male and spouse receipt of nutrition messages by treatment status

| | N | All | Treatment (T) | Control (C) | Normalised difference between (T) and (C) | P-value |
|---|---|---|---|---|---|---|
| Has received automatic text messages with nutrition information in past two years | 2,204 | 0.300 | 0.429 | 0.171 | 0.587 | 0.000 |
| | | (0.459) | (0.495) | (0.377) | | |
| Doesn't know if receives automatic text messages | 2,204 | 0.010 | 0.009 | 0.012 | -0.027 | 0.550 |
| | | (0.102) | (0.095) | (0.108) | | |
| Receives nutrition messages from government hospital | 2,204 | 0.005 | 0.005 | 0.004 | 0.027 | 0.519 |
| | | (0.067) | (0.074) | (0.060) | | |
| Receives nutrition messages from NGO/international organisation | 2,204 | 0.031 | 0.044 | 0.018 | 0.151 | 0.005 |
| | | (0.174) | (0.206) | (0.134) | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| Receives nutrition messages from church/mosque | 2,204 | 0.000 | 0.001 | 0.000 | 0.043 | 0.317 |
| | | (0.021) | (0.030) | (0.000) | | |
| Receives nutrition messages from Wazazi Nipendeni/15501 | 2,204 | 0.162 | 0.245 | 0.079 | 0.461 | 0.000 |
| | | (0.369) | (0.430) | (0.270) | | |
| Receives nutrition messages from 15014 number | 2,204 | 0.020 | 0.033 | 0.006 | 0.191 | 0.000 |
| | | (0.138) | (0.178) | (0.080) | | |
| Receives nutrition messages from an unknown source | 2,204 | 0.102 | 0.137 | 0.066 | 0.235 | 0.000 |
| | | (0.302) | (0.344) | (0.249) | | |
| Spouse has received automatic text messages with nutrition information in past two years | 2,113 | 0.117 | 0.185 | 0.050 | 0.430 | 0.000 |
| | | (0.321) | (0.389) | (0.218) | | |
| Doesn't know if spouse receives automatic text messages with nutrition information | 2,113 | 0.291 | 0.298 | 0.285 | 0.029 | 0.578 |
| | | (0.454) | (0.457) | (0.451) | | |
| Spouse receives nutrition messages from government hospital | 2,113 | 0.000 | 0.001 | 0.000 | 0.044 | 0.317 |
| | | (0.022) | (0.031) | (0.000) | | |
| Spouse receives nutrition messages from NGO/international organisation | 2,113 | 0.006 | 0.010 | 0.002 | 0.110 | 0.023 |
| | | (0.078) | (0.102) | (0.043) | | |
| Spouse receives nutrition messages from church/mosque | 2,113 | 0.000 | 0.000 | 0.000 | | |
| | | (0.000) | (0.000) | (0.000) | | |
| Spouse receives nutrition messages from Wazazi Nipendeni/15501 | 2,113 | 0.064 | 0.107 | 0.023 | 0.348 | 0.000 |
| | | (0.245) | (0.309) | (0.148) | | |
| Spouse receives nutrition messages from 15014 number | 2,113 | 0.008 | 0.013 | 0.003 | 0.118 | 0.011 |
| | | (0.089) | (0.115) | (0.053) | | |
| Spouse receives nutrition messages from unknown source | 2,113 | 0.045 | 0.069 | 0.022 | 0.228 | 0.000 |
| | | (0.207) | (0.253) | (0.145) | | |
| Has seen messages that spouse received from Wazazi Nipendeni | 103 | 0.864 | 0.885 | 0.750 | 0.347 | 0.249 |
| | | (0.344) | (0.321) | (0.447) | | |

Notes: Estimates from the mNutrition Tanzania endline survey sample. Standard deviations are in parentheses. P-values are from the tests of difference of means between treatment and control group

Source: Authors' own

## C.2 ITT/LATE impact estimates

### Table 10.3: Impact estimates of impact of mNutrition on female IYCF knowledge, individual indicators

| | Control mean | Impact estimates, basic controls | Impact estimates, extended controls | N |
|---|---|---|---|---|
| When should a baby start getting breast milk? | 0.869 | 0.018 | 0.017 | 2,469 |
| | | (0.014) | (0.014) | |
| What should a mother do with colostrum? | 0.833 | 0.008 | 0.004 | 2,469 |
| | | (0.017) | (0.016) | |
| When should a baby be breastfed? | 0.712 | 0.002 | 0.002 | 2,469 |
| | | (0.025) | (0.024) | |
| What should a mother do if her baby is not getting enough breast milk? | 0.775 | 0.006 | 0.007 | 2,469 |
| | | (0.018) | (0.018) | |
| Should a baby be given water during hot weather? | 0.699 | -0.006 | -0.009 | 2,469 |
| | | (0.021) | (0.020) | |
| At what age should babies be given water or other liquids? | 0.739 | -0.012 | -0.008 | 2,469 |
| | | (0.019) | (0.019) | |
| At what age should babies be given other foods? | 0.826 | -0.041** | -0.036** | 2,469 |
| | | (0.018) | (0.016) | |
| Name one thing that can happen to a child if they do not get enough iron | 0.352 | 0.039* | 0.035* | 2,469 |
| | | (0.020) | (0.018) | |
| Name foods rich in vitamin A | 0.765 | -0.020 | -0.025 | 2,469 |
| | | (0.021) | (0.021) | |
| Name foods rich in iron | 0.479 | 0.018 | 0.014 | 2,469 |
| | | (0.026) | (0.024) | |
| What are strategies to protect a child from intestinal worms? | 0.357 | -0.012 | -0.017 | 2,469 |
| | | (0.023) | (0.022) | |
| How many times should non-breastfed children aged 6–23 months be fed each day? | 0.476 | 0.042 | 0.033 | 2,469 |
| | | (0.030) | (0.027) | |
| Until what age (in months) should children continue to be breastfed? | 0.769 | 0.008 | 0.015 | 2,469 |
| | | (0.020) | (0.019) | |
| Should children over 12 months be given deworming medication? | 0.857 | 0.030** | 0.027* | 2,469 |
| | | (0.015) | (0.014) | |
| Do breastfeeding mothers need extra calories and more nutritious food than non-breastfeeding mothers? | 0.850 | 0.015 | 0.011 | 2,469 |
| | | (0.014) | (0.014) | |
| Should you wash utensils before preparing and storing food? | 0.998 | -0.003 | -0.003 | 2,469 |
| | | (0.002) | (0.002) | |
| How should you wash utensils? | 0.845 | 0.008 | 0.009 | 2,469 |
| | | (0.016) | (0.016) | |
| How does giving animal-sourced foods to children aged 6–23 months affect their growth? | 0.937 | -0.007 | -0.006 | 2,469 |
| | | (0.011) | (0.010) | |
| | 0.489 | 0.028 | 0.024 | 2,469 |

| | | | | |
|---|---|---|---|---|
| Is it important for a mother to take iron tablets after giving birth? | | (0.025) | (0.025) | |
| For how many months after giving birth should a mother continue to take iron tablets? | 0.166 | 0.038** | 0.036** | 2,469 |
| | | (0.017) | (0.017) | |

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 the treatment from an OLS regression of the outcome of interest on the 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, age of individual responding, whether individual responding is literate in Swahili, PPI score, and whether individual owns a mobile phone. Control mean is comparison group's mean at endline. * p<0.10 **p<0.05 ***p<0.01

Source: Authors' own

### Table 10.4: Impact estimates of impact of mNutrition on male IYCF knowledge, individual indicators

| | Control mean | Impact estimates, basic controls | Impact estimates, extended controls | N |
|---|---|---|---|---|
| When should a baby start getting breast milk? | 0.602 | 0.034 | 0.028 | 2,204 |
| | | (0.023) | (0.022) | |
| What should a mother do with colostrum? | 0.470 | 0.039 | 0.026 | 2,204 |
| | | (0.025) | (0.023) | |
| When should a baby be breastfed? | 0.503 | 0.008 | 0.007 | 2,204 |
| | | (0.028) | (0.028) | |
| What should a mother do if her baby is not getting enough breast milk? | 0.729 | 0.012 | 0.014 | 2,204 |
| | | (0.022) | (0.023) | |
| Should a baby be given water during hot weather? | 0.540 | 0.041* | 0.035 | 2,204 |
| | | (0.023) | (0.022) | |
| At what age should babies be given water or other liquids? | 0.537 | 0.061** | 0.058** | 2,204 |
| | | (0.025) | (0.025) | |
| At what age should babies be given other foods? | 0.673 | -0.011 | -0.010 | 2,204 |
| | | (0.022) | (0.021) | |
| Name one thing that can happen to a child if they do not get enough iron? | 0.345 | 0.032 | 0.030 | 2,204 |
| | | (0.023) | (0.020) | |
| Name foods rich in vitamin A | 0.698 | 0.030 | 0.026 | 2,204 |
| | | (0.023) | (0.023) | |
| Name foods rich in iron | 0.478 | 0.029 | 0.025 | 2,204 |
| | | (0.028) | (0.027) | |
| What are strategies to protect a child from intestinal worms? | 0.352 | 0.013 | 0.016 | 2,204 |
| | | (0.023) | (0.021) | |
| How many times should non-breastfed children aged 6–23 months be fed each day? | 0.320 | 0.008 | 0.001 | 2,204 |
| | | (0.025) | (0.023) | |
| Until what age (in months) should children continue to be breastfed? | 0.664 | -0.004 | -0.001 | 2,204 |
| | | (0.021) | (0.021) | |
| Should children over 12 months be given deworming medication? | 0.748 | 0.006 | 0.007 | 2,204 |
| | | (0.019) | (0.019) | |

| | | | | |
|---|---|---|---|---|
| Do breastfeeding mothers need extra calories and more nutritious food than non-breastfeeding mothers? | 0.855 | 0.022 | 0.021 | 2,204 |
| | | (0.016) | (0.016) | |
| Should you wash utensils before preparing and storing food? | 0.990 | 0.001 | 0.002 | 2,204 |
| | | (0.004) | (0.005) | |
| How should you wash utensils? | 0.837 | -0.015 | -0.010 | 2,204 |
| | | (0.016) | (0.016) | |
| How does giving animal-sourced foods to children aged 6–23 months affect their growth? | 0.933 | 0.008 | 0.010 | 2,204 |
| | | (0.011) | (0.011) | |
| Is it important for a mother to take iron tablets after giving birth? | 0.403 | -0.007 | -0.001 | 2,204 |
| | | (0.024) | (0.021) | |
| For how many months after giving birth should a mother continue to take iron tablets? | 0.115 | 0.011 | 0.014 | 2,204 |
| | | (0.014) | (0.014) | |

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 the treatment from an OLS regression of the outcome of interest on the 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, age of individual responding, whether individual responding is literate in Swahili, PPI score, and whether individual owns a mobile phone. Control mean is comparison group's mean at endline. * p<0.10 **p<0.05 ***p<0.01

Source: Authors' own

### Table 10.5: LATE of impact of mNutrition use on females' IYCF knowledge, individual answers

| | Control mean | LATE estimates, basic controls | LATE estimates, extended controls | N |
|---|---|---|---|---|
| When should a baby start getting breast milk? | 0.869 | 0.045 | 0.040 | 2,469 |
| | | (0.035) | (0.034) | |
| What should a mother do with colostrum? | 0.833 | 0.020 | 0.008 | 2,469 |
| | | (0.041) | (0.040) | |
| When should a baby be breastfed? | 0.712 | 0.005 | 0.012 | 2,469 |
| | | (0.062) | (0.062) | |
| What should a mother do if her baby is not getting enough breast milk? | 0.775 | 0.015 | 0.019 | 2,469 |
| | | (0.043) | (0.044) | |
| Should a baby be given water during hot weather? | 0.699 | -0.014 | -0.023 | 2,469 |
| | | (0.052) | (0.051) | |
| At what age should babies be given water or other liquids? | 0.739 | -0.029 | -0.019 | 2,469 |
| | | (0.046) | (0.047) | |
| At what age should babies be given other foods? | 0.826 | -0.101** | -0.090** | 2,469 |
| | | (0.042) | (0.041) | |
| Name one thing that can happen to a child if they do not get enough iron | 0.352 | 0.095* | 0.088* | 2,469 |
| | | (0.049) | (0.045) | |
| Name foods rich in vitamin A | 0.765 | -0.050 | -0.063 | 2,469 |
| | | (0.051) | (0.052) | |
| Name foods rich in iron | 0.479 | 0.045 | 0.033 | 2,469 |

| | | | | |
|---|---|---|---|---|
| | | (0.062) | (0.061) | |
| What are strategies to protect a child from intestinal worms? | 0.357 | -0.030 | -0.041 | 2,469 |
| | | (0.057) | (0.055) | |
| How many times should non-breastfed children aged 6–23 months be fed each day? | 0.476 | 0.102 | 0.092 | 2,469 |
| | | (0.073) | (0.069) | |
| Until what age (in months) should children continue to be breastfed? | 0.769 | 0.019 | 0.039 | 2,469 |
| | | (0.049) | (0.047) | |
| Should children over 12 months be given deworming medication? | 0.857 | 0.074** | 0.074** | 2,469 |
| | | (0.035) | (0.036) | |
| Do breastfeeding mothers need extra calories and more nutritious food than non-breastfeeding mothers? | 0.850 | 0.038 | 0.033 | 2,469 |
| | | (0.034) | (0.035) | |
| Should you wash utensils before preparing and storing food? | 0.998 | -0.006 | -0.007 | 2,469 |
| | | (0.006) | (0.006) | |
| How should you wash utensils? | 0.845 | 0.019 | 0.029 | 2,469 |
| | | (0.039) | (0.040) | |
| How does giving animal-sourced foods to children aged 6–23 months affect their growth? | 0.937 | -0.017 | -0.008 | 2,469 |
| | | (0.028) | (0.027) | |
| Is it important for a mother to take iron tablets after giving birth? | 0.489 | 0.068 | 0.063 | 2,469 |
| | | (0.062) | (0.062) | |
| For how many months after giving birth should a mother continue to take iron tablets? | 0.166 | 0.092** | 0.089** | 2,469 |
| | | (0.041) | (0.042) | |

Notes: Estimates from the mNutrition Tanzania endline survey sample. Standard errors are in parentheses and clustered at the village level. Reported are the second-stage estimates from the instrumental variable regression, where an indicator variable of household-level mNutrition use over the last two years is instrumented by the 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, age of individual responding, whether individual responding is literate in Swahili, PPI score, and whether individual owns a mobile phone. Control mean is comparison group's mean at endline. * p<0.10 **p<0.05 ***p<0.01

Source: Authors' own

### Table 10.6: LATE of impact of mNutrition use on males' IYCF knowledge, individual answers

| | Control mean | LATE estimates, basic controls | LATE estimates, extended controls | N |
|---|---|---|---|---|
| When should a baby start getting breast milk? | 0.602 | 0.089 | 0.075 | 2,204 |
| | | (0.060) | (0.059) | |
| What should a mother do with colostrum? | 0.470 | 0.101 | 0.070 | 2,204 |
| | | (0.063) | (0.062) | |
| When should a baby be breastfed? | 0.503 | 0.021 | 0.018 | 2,204 |
| | | (0.074) | (0.075) | |
| What should a mother do if her baby is not getting enough breast milk? | 0.729 | 0.031 | 0.037 | 2,204 |
| | | (0.059) | (0.061) | |
| Should a baby be given water during hot weather? | 0.540 | 0.107* | 0.092 | 2,204 |
| | | (0.059) | (0.059) | |
| | 0.537 | 0.161** | 0.154** | 2,204 |

| | | | | |
|---|---|---|---|---|
| At what age should babies be given water or other liquids? | | (0.067) | (0.068) | |
| At what age should babies be given other foods? | 0.673 | -0.028 | -0.026 | 2,204 |
| | | (0.057) | (0.056) | |
| Name one thing that can happen to a child if they do not get enough iron | 0.345 | 0.085 | 0.080 | 2,204 |
| | | (0.059) | (0.054) | |
| Name foods rich in vitamin A | 0.698 | 0.078 | 0.070 | 2,204 |
| | | (0.060) | (0.061) | |
| Name foods rich in iron | 0.478 | 0.077 | 0.067 | 2,204 |
| | | (0.072) | (0.073) | |
| What are strategies to protect a child from intestinal worms? | 0.352 | 0.035 | 0.043 | 2,204 |
| | | (0.060) | (0.057) | |
| How many times should non-breastfed children aged 6–23 months be fed each day? | 0.320 | 0.021 | 0.002 | 2,204 |
| | | (0.064) | (0.061) | |
| Until what age (in months) should children continue to be breastfed? | 0.664 | -0.009 | -0.002 | 2,204 |
| | | (0.056) | (0.056) | |
| Should children over 12 months be given deworming medication? | 0.748 | 0.016 | 0.019 | 2,204 |
| | | (0.049) | (0.051) | |
| Do breastfeeding mothers need extra calories and more nutritious food than non-breastfeeding mothers? | 0.855 | 0.059 | 0.056 | 2,204 |
| | | (0.041) | (0.042) | |
| Should you wash utensils before preparing and storing food? | 0.990 | 0.002 | 0.004 | 2,204 |
| | | (0.012) | (0.012) | |
| How should you wash utensils? | 0.837 | -0.039 | -0.026 | 2,204 |
| | | (0.042) | (0.041) | |
| How does giving animal-sourced foods to children aged 6–23 months affect their growth? | 0.933 | 0.022 | 0.026 | 2,204 |
| | | (0.030) | (0.029) | |
| Is it important for a mother to take iron tablets after giving birth? | 0.403 | -0.018 | -0.004 | 2,204 |
| | | (0.062) | (0.057) | |
| For how many months after giving birth should a mother continue to take iron tablets? | 0.115 | 0.030 | 0.036 | 2,204 |
| | | (0.038) | (0.038) | |

Notes: Estimates from the mNutrition Tanzania endline survey sample. Standard errors are in parentheses and clustered at the village level. Reported are the second-stage estimates from the instrumental variable regression, where an indicator variable of household-level mNutrition use over the last two years is instrumented by the 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, and PPI score. Control mean is comparison group's mean at endline. * p<0.10 **p<0.05 ***p<0.01

Source: Authors' own

## C.3 Robustness of impact estimates

**Table 10.7: P-values for child-level outcomes based on cluster-robust standard errors and randomisation inference**

| | N | P-value from cluster-robust standard errors | P-value from randomisation inference | Standard error for RI-based P-value |
|---|---|---|---|---|
| Child HAZ | 2,794 | 0.270 | 0.266 | 0.0045 |
| Child WHZ | 2,792 | 0.592 | 0.596 | 0.0049 |
| Number of food groups (of seven) children aged 6–35 months consume | 2,558 | 0.073 | 0.080 | 0.0026 |
| Children aged 6–35 months who consume four or more food groups | 2,558 | 0.093 | 0.091 | 0.0028 |
| Children aged 6–23 months who meet the MAD | 830 | 0.042 | 0.041 | 0.0019 |
| Children aged 6–23 months who meet the minimum meal frequency | 846 | 0.343 | 0.357 | 0.0048 |

Notes: Estimates from the mNutrition Tanzania endline survey sample. P-values from cluster-robust standard errors correspond to the p-values in the main ITT tables, which allow for any arbitrary within-village correlation in the error terms. P-values from randomisation inference are based on 10,000 repetitions and are for the sharp null hypothesis of no treatment effect for any observation. The standard errors for the randomisation inference-based p-values are derived from the number of repetitions run and the basic specification used to estimate the impacts.

Source: Authors' own

### Table 10.8: P-values for household-level outcomes based on cluster-robust standard errors and randomisation inference

| | N | P-value from cluster-robust standard errors | P-value from randomisation inference | Standard error for RI-based P-value |
|---|---|---|---|---|
| WDDS (1–10) | 2,535 | 0.279 | 0.280 | 0.0045 |
| Met MDD-W | 2,546 | 0.062 | 0.060 | 0.0024 |
| Percentage of correct IYCF knowledge answers (male) | 2,469 | 0.229 | 0.227 | 0.0042 |
| Percentage of correct IYCF knowledge answers (female) | 2,204 | 0.019 | 0.018 | 0.0013 |

Notes: Estimates from the mNutrition Tanzania endline survey sample. P-values from cluster-robust standard errors correspond to the p-values in the main ITT tables, which allow for any arbitrary within-village correlation in the error terms. P-values from randomisation inference are based on 10,000 repetitions and are for the sharp null hypothesis of no treatment effect for any observation. The standard errors for the randomisation inference-based p-values are derived from the number of repetitions run and the basic specification used to estimate the impacts.

Source: Authors' own

### Table 10.9: Impact estimates of impact of mNutrition on main outcomes, with weights to adjust for attrition

| | Control mean | Unweighted impact estimates, basic controls | Weighted impact estimates, basic controls | N |
|---|---|---|---|---|
| Child HAZ | -1.976 | -0.060 | -0.061 | 2,794 |
| | | (0.055) | (0.055) | |
| Child WHZ | 0.182 | 0.029 | 0.030 | 2,792 |
| | | (0.052) | (0.052) | |
| Number of food groups (of seven) children aged 6–35 months consume | 3.766 | 0.110* | 0.109* | 2,558 |
| | | (0.060) | (0.060) | |
| Children aged 6–35 months who consume four or more food groups | 0.604 | 0.039* | 0.039* | 2,558 |
| | | (0.023) | (0.023) | |

| | | | | |
|---|---|---|---|---|
| Children aged 6–23 months who meet the MAD | 0.224 | 0.071** | 0.072** | 830 |
| | | (0.034) | (0.034) | |
| Children aged 6–23 months who meet the minimum meal frequency | 0.790 | 0.025 | 0.025 | 846 |
| | | (0.026) | (0.026) | |
| WDDS (1–10) | 4.437 | 0.073 | 0.069 | 2,535 |
| | | (0.068) | (0.068) | |
| Met MDD-W | 0.625 | 0.040* | 0.038* | 2,546 |
| | | (0.021) | (0.021) | |
| Percentage of correct answers (female) | 68.969 | 0.795 | 0.833 | 2,469 |
| | | (0.659) | (0.659) | |
| Percentage of correct answers (male) | 58.959 | 1.772** | 1.775** | 1,210 |
| | | (0.896) | (0.899) | |

Notes: Estimates from the mNutrition Tanzania endline survey sample. Standard errors are in parentheses and clustered at the village level. Impact estimates in the second column are coefficients on treatment from an unweighted OLS regression of the outcome of interest on the treatment variable. Impact estimates in the third column come from otherwise identical weighted OLS regressions, where the inverse of the predicted probability of endline survey completion is used as a sampling weight. The probability is predicted through logit regression of completion status on covariates collected at baseline: household size, whether household head is female, whether household head is literate in Swahili, whether primary female owns a mobile phone, and PPI score. Models in both columns control for baseline household classification and value of the respective outcome at baseline. Control mean is comparison group's mean at endline. * p<0.10 **p<0.05 ***p<0.01

Source: Authors' own

# Annex D Estimation strategy details

## D.1 ITT estimates

Because the offer of treatment was randomly assigned, we use the systematic variation in the random offer of access to the mNutrition service 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 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—we use ANCOVA to generate our primary estimates, and simple differences and difference-in-differences models as robustness checks.

For outcomes that were not observable at baseline—such as children's dietary diversity, for example—we use a simple differences specification to estimate treatment effects. To identify the simple differences treatment effects, we estimate the following OLS regressions:

$$Y_{1ihv} = \beta_0 + \beta_{1,SD} Treatment_v + \delta_x X_{0ihv} + \varepsilon_{ihv} \,,$$

where $Y_{1ihv}$ is the outcome measured at endline ($t = 1$), for individual $i$, in household $h$, in village $v$, $\beta_0$ is a constant term, $Treatment_v$ is an indicator equal to 1 if village $v$ was randomly assigned to the treatment group, $X_{0ihv}$ is a vector of observable characteristics for individual $i$ measured at baseline ($t = 0$), and $\varepsilon_{ihv}$ is an error term that we cluster at the village level. In this model, $\delta_x$ represents the vector of coefficients on the controls and $\beta_{1,SD}$ is the parameter of interest: the simple differences causal effect of being offered access to the mNutrition service on the outcome.

Though the random assignment of treatment to villages ensures that $\beta_{1,SD}$ is an unbiased estimate of the causal effect of the mNutrition offer, more efficient methods are available for outcomes with both baseline and endline measures. In particular, the difference-in-differences specification can be estimated by running the following OLS regression:

$$Y_{tihv} = \sum_{t=0}^{1} \delta_t + \beta_{1,DD} Treatment_{tv} + \beta_2 EverTreat_v + \delta_x X_{tihv} + \varepsilon_{tihv}$$

where $t \in \{0,1\}$, $Y_{tihv}$ is the outcome measured in period $t$, the $\delta_t$ are time period fixed effects, $EverTreat_v$ is an indicator for whether village $v$ ever receives the mNutrition service offers, $Treatment_{tv}$ is an indicator for whether village $v$ received offers for the mNutrition service in period $t$, $X_{tihv}$ is a vector of time varying controls, and $\varepsilon_{tihv}$ is an error term that, again, is clustered at the village level. For outcomes with an autocorrelation above 0.5, the difference-in-differences estimator will yield more precise estimates of the treatment effects ($\beta_{1,DD}$) than the simple differences model.

However, with one baseline survey and one endline survey, treatment effects estimated through an ANCOVA specification are expected to have lower variance[25] than the difference-in-

[25] Although we know the variance of the treatment effects estimated through ANCOVA will be lower than the variance for those estimated with simple differences or difference-in-differences, we cannot know the size of the difference between the variances until we have collected endline data and estimated the autocorrelations for each outcome.

differences treatment effects as long as the autocorrelation for the outcome is below 1.[26] ANCOVA treatment effects can be estimated by running the following OLS regression:

$$Y_{1ihv} = \beta_0 + \beta_{1,A} Treatment_v + \beta_Y Y_{0ihv} + \delta_x X_{0ihv} + \varepsilon_{ihv}$$

where all of the parameters are defined in the same way as in the above simple differences equation and $Y_{0ihv}$ is the baseline measure of the outcome. 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, we will rely on this ANCOVA specification to generate the main treatment effect estimates.

## D.2 LATE for compliers

The specifications described above enable us 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[27] we can also estimate the LATE of receiving the mNutrition content for compliers, i.e. 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.[28] Both of these assumptions are likely to be satisfied in the context of this evaluation (see the full discussion in Gilligan *et al*., (2018)).

If both assumptions are satisfied and participation in the mNutrition service is reliably observed in the data, the LATE for complier households can be estimated through 2SLS estimates of the ANCOVA or simple difference models discussed in the previous subsection, using the random assignment of villages to receive the mNutrition treatment offer as an excluded instrument for observed take-up of the service.[29] Specifically, we will estimate the following models:

$$Y_{1ihv} = \beta_0 + \beta_{1,2SLS} \widehat{TT_{hv}} + \beta_Y Y_{0ihv} + \delta_x X_{0ihv} + \varepsilon_{ihv}$$

$$TT_{hv} = \gamma_0 + \gamma_{1,A} Treatment_v + \gamma_Y Y_{0ihv} + \gamma_x X_{0ihv} + u_{ihv}$$

where $Treatment_v$ is the indicator for whether the household resides in a village $v$ that was assigned to receive the mNutrition offer, and $TT_{hv}$ is an indicator for whether household $h$ in village $v$ actually registers for the mNutrition service. $\widehat{TT_{hv}}$ is the predicted value for the take-up of household $h$ in village $v$ from the take-up equation (the second equation listed above). In this context, $\beta_{1,2SLS}$ represent the estimated effect of receiving the mNutrition content for the sub-

[26] See McKenzie (2012). In an experiment with one baseline and one follow-up survey, the ratio of the variance of the difference-in-differences estimate to the variance of the ANCOVA variance is given by $\frac{2}{(1+\rho)}$, where ρ is the autocorrelation.
[27] See Imbens and Rubin (2015) for a complete discussion.
[28] 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.
[29] For the linear models specified in this context, the 2SLS estimates of LATE for compliers will be equal to the ratio of the ITT estimate of the impact of the treatment offer on the outcome to the ITT estimate of the impact of the treatment offer on take-up of the programme.

sample of households that are induced to participate in the service by the randomly assigned offer.

The LATE treatment effects for compliers provide a different, but still policy-relevant parameter: they represent, albeit for a specific sub-population (compliers), the causal effect of exposure to the mNutrition messaging. Regardless, in the following sections we estimate and discuss both parameters in order to provide more complete conclusions about the causal effects of the mNutrition service.

## D.3 Randomisation inference

With 180 study clusters included in the quantitative evaluation, statistical inference based on cluster-robust standard errors is likely to be valid and to result in tests of the correct size (Bertrand *et al.*, 2004). However, for each outcome we also conduct randomisation inference as a robustness check on the tests that rely on the asymptotic normality of a test statistic based on a finite sample (Fisher, 1935; Rosenbaum, 2002; Greevy *et al.,* 2004; Imbens and Rosenbaum, 2005; Small *et al*., 2008). Randomisation inference offers a non-parametric alternative for testing the sharp null hypothesis of no treatment effect for any household.[30]

To conduct randomisation inference for the sharp null hypothesis of no treatment effect for any household, we proceed as follows. First, we calculate the test statistic, for example the difference in mean outcomes between villages assigned to the mNutrition treatment group and villages assigned to the mNutrition control group, $\hat{\beta} = \bar{y}_{ihv,T=1} - \bar{y}_{ihv,T=0}$. Next, we conduct $R = 10{,}000$ placebo treatment assignments. That is, we re-assign 90 of the 180 sample villages to a placebo treatment group and the other 90 to the placebo control group, 10,000 different times. For each repetition $r$, we calculate the treatment effect, which under the null of no effect for any household is simply $\hat{\beta}_r = \bar{y}_{ihv,T_r=1} - \bar{y}_{ihv,T_r=0}$. Here $T_r = 1$ denotes that the household resides in a village that was assigned to the placebo treatment group in repetition $r$, $T_r = 0$ indicates that the village was assigned to the placebo control group in repetition $r$, and by the null of no effect for any household $y_{ihv,T_r=1} = y_{ihv,T_r=0}$, so the observed outcome does not need to be adjusted.

After performing the placebo treatment assignment 10,000 times, we will have a distribution of test statistics from all of the repetitions: $\{\hat{\beta}_r\} = \{\hat{\beta}_1, \dots, \hat{\beta}_R\}$. To assess the plausibility of the observed test statistic ($\hat{\beta}$) under the null hypothesis, we calculate the share of repetitions for which $\hat{\beta}_r > \hat{\beta}$. This share is an empirical p-value for the sharp null of no treatment effect for any household.[31]

[30] In practice, randomisation inference can be used to test the sharp null of a treatment effect of any size, not just no treatment effect.

[31] Note that, although the exact p-value could be computed by generating all $\binom{180}{90}$ possible treatment assignments, this is computationally infeasible. With 10,000 repetitions, the empirical p-value will have a standard error less than or equal to $\frac{1}{2\sqrt{R}} = 0.005$.

# Annex E Endline questionnaire outline

**Study on mobile phone technology based nutrition and health advisory services in Tanzania**
**ENDLINE SURVEY: Household Questionnaire – 2 September 2018**

*Outline:*

**Module A: Household Front Page Identification**
**Module B: Household Composition and Education**
Part 1: Household Roster
**Module C: Housing and Assets**
Part 1: Housing
**Module D: General Health**
Part 1: Health and Vaccinations for Those Under Five
Part 2: HIV/AIDS Awareness
**Module E: Marriage and Fertility History**
Part 1: Marriage History (Female)
Part 2a: Desired Fertility Preferences (Female)
Part 2b: Desired Fertility Preferences (Male)
**Module F: Mobile Phone Access and Usage**
Part 1a: Mobile Phone Access and Usage (Female)
Part 1b: Mobile Phone Access and Usage (Male)
Part 2: Household Mobile Usage and Access
**Module G: Antenatal and Postnatal Care**
**Module H: IYCF Knowledge and Beliefs**
Part 1: IYCF Knowledge and Beliefs (Female)
Part 2: IYCF Knowledge and Beliefs (Male)
**Module I: Nutrition Practices**
Part 1: Infant and Young Child Feeding (IYCF) Practices
Part 2: Dietary Diversity
**Module M: Non-Food Expenditure**
Part 1: Non-Food Expenditure (Monthly)
Part 2: Non-Food Expenditure (Yearly)
**Module N: Food Consumption and Expenditure**
Part 1: Food Consumption in the Last Week
**Module J: Trust Likelihood of Nutrition and Health Information**
**Module L: Interaction with the Programme**
Part 1a: Wazazi Nipendeni Programme (Female)
Part 2a: Use of the Wazazi Nipendeni Programme (Female)
Part 3a: Quality of the Wazazi Nipendeni Programme (Female)
Part 1b: Wazazi Nipendeni Programme (Female)
Part 2b: Use of the Wazazi Nipendeni Programme (Male)
Part 3b: Quality of the Wazazi Nipendeni Programme (Male)
**Module Q: Convex Time Budgets Elicitation**
Part 1e: Examples
Part 1: Game 1
Part 2e: Extra Questions
Part 2: Game 2
**Module K: Anthropometry**