Evaluation and Program Planning 73 (2019) 1–9

Contents lists available at ScienceDirect

Evaluation and Program Planning

journal homepage: www.elsevier.com/locate/evalprogplan

Monitoring and evaluation design of Malawi's Right Foods at the Right Time nutrition program

Julie C. Ruel-Bergeron[a], Kristen M. Hurley[a,*], Yunhee Kang[a], Nancy Aburto[b], Arghanoon Farhikhtah[b], Alessandro Dinucci[b], Luca Molinas[b], Lee Shu Fune Wu[a], Maithilee Mitra[a], John Phuka[c], Rolf Klemm[a,d], Keith West[a], Parul Christian[a,e]

[a] Johns Hopkins Bloomberg School of Public Health, Department of International Health, Program in Human Nutrition, 615 N. Wolfe St., Baltimore, MD, 21205, USA
[b] United Nations World Food Programme, Nutrition Division, Via Cesare Guilio Viola 68, Parco dei Medici, Rome, 00148, Italy
[c] College of Medicine, University of Malawi, Private Bag 360, Blantyre, Malawi
[d] Helen Keller International, 352 Park Avenue South, 12th floor, New York, NY, 10010, USA
[e] The Bill & Melinda Gates Foundation, Women's Nutrition, 500 Fifth Avenue North, Seattle, WA, 98109, USA

ARTICLE INFO

Keywords:
Monitoring
Evaluation
Nutrition
Stunting
Malawi

ABSTRACT

Child stunting is a public health problem in Malawi. In 2014, the Government of Malawi launched the Right Foods at the Right Time (RFRT) program in Ntchisi district delivering nutrition social and behavior change communication, a small-quantity lipid-based nutrient supplement to children 6–23 months, and nutrition sensitive activities. Monitoring and evaluation (M&E) systems are key aspects of successful program implementation. We describe these and the methodology for an impact evaluation that was conducted for this program. Two monitoring systems using traditional and electronic platforms were established to register and track program delivery and processes including number of eligible beneficiaries, worker performance, program participation, and to monitor input, output, and outcome indicators. The impact evaluation used comparative cross-sectional and longitudinal designs to assess impact on anthropometric and infant and young child feeding outcomes. Three cross-sectional surveys (base-, mid-, and end-line) and two longitudinal cohorts of children followed in 6-month intervals from 6 to 24 months of age, were conducted in sampled households in the program and a neighboring comparison district. Additional M&E included qualitative studies, a process evaluation, and a cost-effectiveness study. The current paper describes lessons from this program's M&E, and demonstrates how multiple implementation research activities can inform course-correction and program scale-up.

1. Introduction

Stunting, defined as length-for-age z-score < -2, affects 37% of all children under five years of age in Malawi (National Statistical Office, ICF Macro, 2017). The consequences of stunting are significant and contribute to increased childhood morbidity, as well as poor physical and cognitive development (Black, Victora, & Walker, 2013; Prendergast & Humphrey, 2014). Further, the fragility of the immune system of a stunted child contributes to increased mortality, even in mild cases of stunting (Olofin, McDonald, & Ezzati, 2013). As Olofin and colleagues have demonstrated, children with a height-for-age z-score (HAZ) between -1 and -2 have an increased risk of mortality of 1.56 (95% CI: 0.98, 2.46), while those with a HAZ between -2 and -3 have an increased risk of mortality of 2.45 (95% CI: 1.56, 3.87) as compared to those with a HAZ ≥-1 (Olofin et al., 2013).

The determinants of child stunting in Malawi and similar contexts are multi-layered, multi-causal, and multi-generational, as outlined in the conceptual framework for maternal and child undernutrition (Black, Allen, & Bhutta, 2008; United Nations Children's Fund (UNICEF) (1990). Malawi is among the poorest countries in the world, ranking 170 out of 188 countries on the Human Development Index (United Nations Development Programme (UNDP) (2016)), and with a gross national income (GNI) per capita of $340 in 2015 (The World Bank, 2017). Pervasive and chronic poverty, along with a complex and wide range of other socioeconomic factors, are associated with a high prevalence of undernutrition in Malawi (Chikhungu & Madise, 2014; Ruel & Alderman, 2013).

Globally, inappropriate complementary feeding practices have been shown to be associated with stunting (Black et al., 2013); in Malawi, this relationship has also been found to be true (Espo et al., 2002).

* Corresponding author.
E-mail address: khurley2@jhu.edu (K.M. Hurley).

https://doi.org/10.1016/j.evalprogplan.2018.11.001
Received 27 February 2018; Received in revised form 19 October 2018; Accepted 6 November 2018
Available online 11 November 2018

Overall, the diets of young children in much of Malawi are inadequate with only 8% of all children aged 6–23 months consuming a minimum acceptable diet (MAD) (National Statistical Office, ICF Macro, 2017). This low value is largely due to insufficient meal frequency (29% of children achieving a minimum meal frequency (MMF) defined as two meals per day for breastfed children 6–8 months, three meals per day for breastfed children 9–23 months, and four meals per day for non-breastfed children) and limited dietary diversity (only 25% of children 6–23 months consuming the minimum dietary diversity (MDD) of four out of seven food groups in the previous 24 h) (National Statistical Office, ICF Macro, 2017; The World Health Organization (WHO) et al. (2008)).

Malawi was among the first countries to join the Scaling Up Nutrition (SUN) Movement in 2011, and has since demonstrated significant political commitment to addressing child undernutrition (Scaling Up Nutrition, 2011; Pensulo Phiri, 2016). The operationalization of Malawi's nutrition-oriented policies into programming was exemplified in the Government-led nutrition program, 'Right Foods at the Right Time: Targeting Nutrition of Children under Two' (RFRT) (The World Food Programme (WFP), 2015), which was implemented in Ntchisi District of central Malawi. The primary objective of the program was to prevent stunting during the critical window from conception to two years of age, also referred to as the first 1000 days, through a package of evidence-based health and nutrition interventions that targeted pregnant women and children 6–23 months of age. Secondary objectives were related to improvements in other nutritional outcomes, including child underweight, wasting, and anemia, and infant and young child feeding and hygiene behaviors. The program also aimed to influence the global scale-up of prevention programming, while documenting best practices, models, and evidence of impact to enable governments and influence the global nutrition community.

Significant financial resources, intellectual efforts, and partnership-building between the Government of Malawi (GoM), the World Food Programme (WFP), World Vision Malawi (WVM), the Children's Investment Fund Foundation (CIFF), and Johns Hopkins Bloomberg School of Public Health (JHSPH), were invested into the program's design, implementation, funding, and evaluation, respectively. Implementation research activities conducted under the program's monitoring and evaluation (M&E) plan and an independent impact evaluation study were used to generate immediate and longer-term program feedback and evidence, which were used for mid-course adjustments, correction, and continued learning. Furthermore, the M&E plan was crucial for responding to the objectives of the RFRT. The objective of the current manuscript is to describe the development and implementation of the M&E systems and research, as well as lessons learned throughout the process of implementation of each of these program evaluation components.

2. The Right Foods at the Right Time

Following an extensive planning and mass registration phase in 2013, the RFRT program was launched in January 2014 by the GoM with technical, implementation, and financial support from WFP, WVM, and CIFF, respectively. The GoM selected the district of Ntchisi as the program area based on the following criteria: (1) government priority; (2) high burden of stunting; (3) partnership potential; and (4) feasibility of scale-up (The World Food Programme (WFP), 2015). The RFRT program was built on a robust M&E system that was designed following a stepwise approach that provided a continuous flow of information which enabled the program to be adapted and hence respond to identified and changing needs of the population throughout the implementation period (Fig. 1). The details of the program intervention are published elsewhere (The World Food Programme (WFP), 2015). In summary, the originally conceived package included: (1) a social behavior change communication campaign (SBCC) aimed at improving maternal diets, infant and young child feeding (IYCF), and water, sanitation and hygiene (WASH) knowledge and practices; (2) blanket distribution of a daily, small-quantity (20 g) lipid-based nutrient supplement (SQ-LNS), Nutributter (Nutriset, Malaunay, France), to all children 6–23 months of age; and (3) scaling up of community-based management of acute malnutrition (CMAM) to 100% of the district (The World Food Programme (WFP), 2015). A fourth component consisting of nutrition-sensitive actions such as promotion of small livestock and home gardens was incorporated in the second year (2015) of the program. This component aimed at responding to identified challenges in accessing diversified diets in the program district (The World Food Programme (WFP), 2015). The theory of change for these activities and anticipated outcomes are presented in the program logical framework (Fig. 2).

3. Program monitoring systems for the Right Foods at the Right Time

The RFRT's monitoring system was designed with technical support from WFP during the program planning process. The objective of the monitoring system was to monitor process, output, and outcome indicators, as presented in the program's logical framework (Fig. 2), that allowed for quick-response course corrections, or "adaptive programming", and could result in improved program implementation. Three separate yet complementary systems were developed for program monitoring and are described in detail below, including: (1) SCOPE for electronic participant registration and tracking by WFP; (2) Post Distribution Monitoring (PDM) surveys, which were conducted three times per year (January, May and September) and led by WFP; and (3) monthly performance monitoring of output indicators reported by the cooperating implementation partner, WVM (Fig. 3).

3.1. Participant tracking and registration through SCOPE

SCOPE is a WFP-developed electronic beneficiary registration and tracking system for program management (The World Food Programme (WFP), 2014). During a 3-month-long mass registration and mobilization phase in late 2013, children aged 0–23 months were registered into the program at the health center or extended distribution point (EDP) by providing personal details, such as their and their child's names, age or date of birth, village, and what health center they referred to. Once the program was launched, registration was done at monthly distributions, with the aim of registering children at or by the age of six months, when they became eligible to receive the program benefits. Diverse methods of community mobilization (billboards, posters, gatherings, events) and radio publicizing continued throughout the duration of the program to drive continued registration as children became eligible. A plastic beneficiary card with a participant photograph and barcode, that contained the unique SCOPE ID linked their personal details to the program database, was provided to all registered beneficiaries. Beneficiaries presented their cards each month to avail themselves of program benefits, namely receiving the monthly ration of SQ-LNS.

In 2016, the SCOPE system was further streamlined with the development of Mobile SCOPE, which could run on Android version 5.0 mobile devices and was made available at all 12 health centres across the district by 2017. The introduction of the mobile app was in part to respond to feedback received from field monitoring assistants regarding the logistical difficulties of the current system (computers) to register eligible children. Since enrolment rates were already very high across the district (82% mean enrolment from 2014 to 2018), the advantage of mobile registration was rather its improved efficiency for tracing participants, as well as the ability to carry the portable device to remote areas, enabling unscheduled on-site registrations without the need for heavier devices.

During four years of implementation (January 2014-January 2018), SCOPE registered and tracked the attendance of 53,173 beneficiary children between the ages of 6 and 23 months. A key feature of the

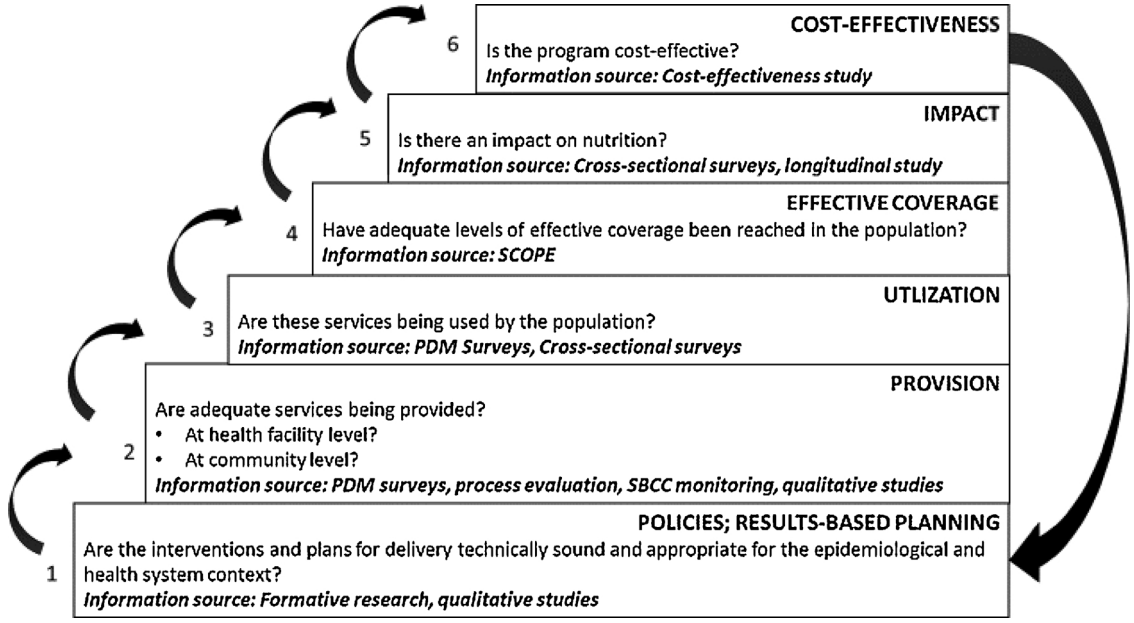

Fig. 1. Flow of information for the RFRT program designed using the stepwise approach to program monitoring and evaluation.
PDM: Post-distribution monitoring; SBCC: Social and behavior chance communication; SCOPE: electronic monitoring system for program coverage.
Source: Adapted from Victora et al., 2010.

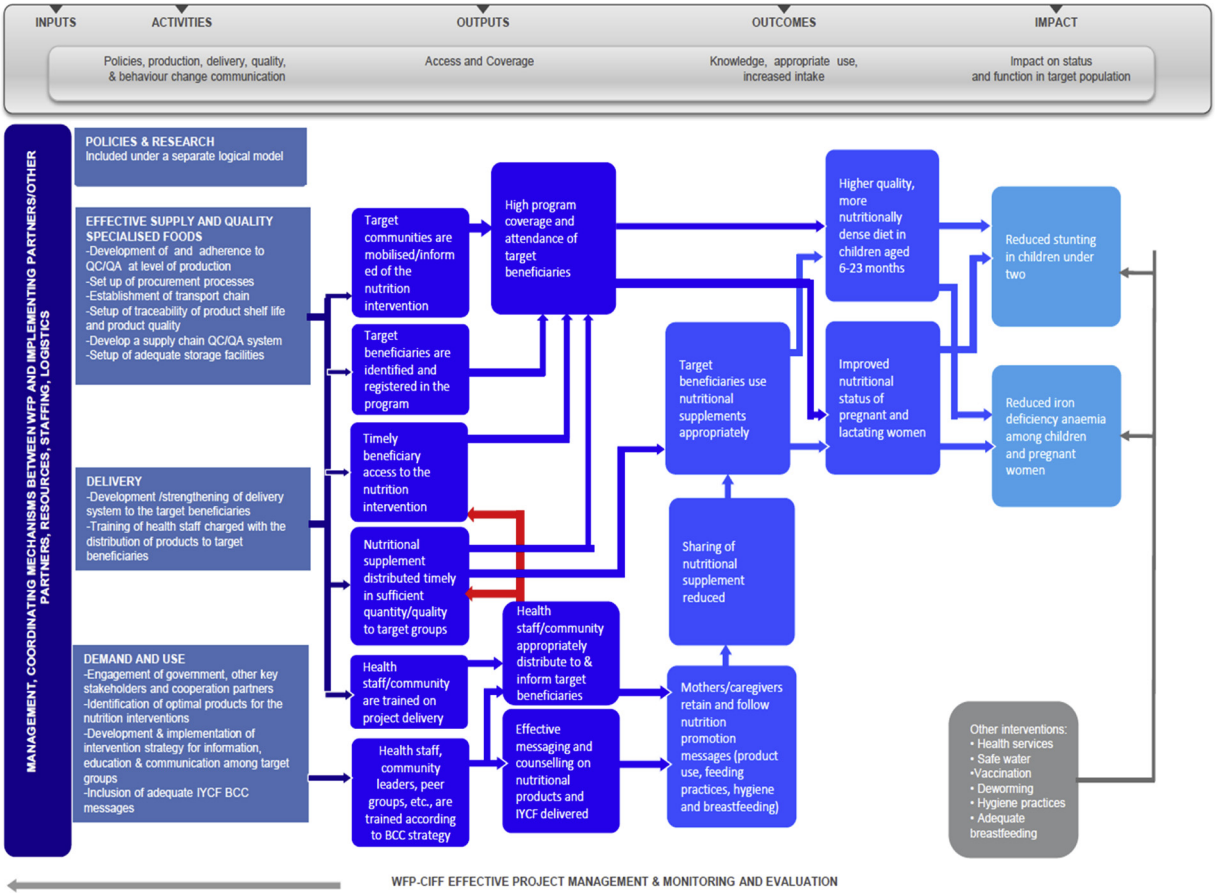

Fig. 2. Logical framework of the RFRT program.

electronic SCOPE system was its functionality in controlling entry and exit criteria to exclude children who were not eligible for the program due to their age (< 6 and ≥24 months). Incoming data from SCOPE could be visualized in a business intelligence dashboard created specifically for this program, which provided near real-time updates on overall participation by village and by monthly distribution cycle. The dashboard also allowed near real-time analysis of participation trends over time and calculated total number of beneficiaries reached

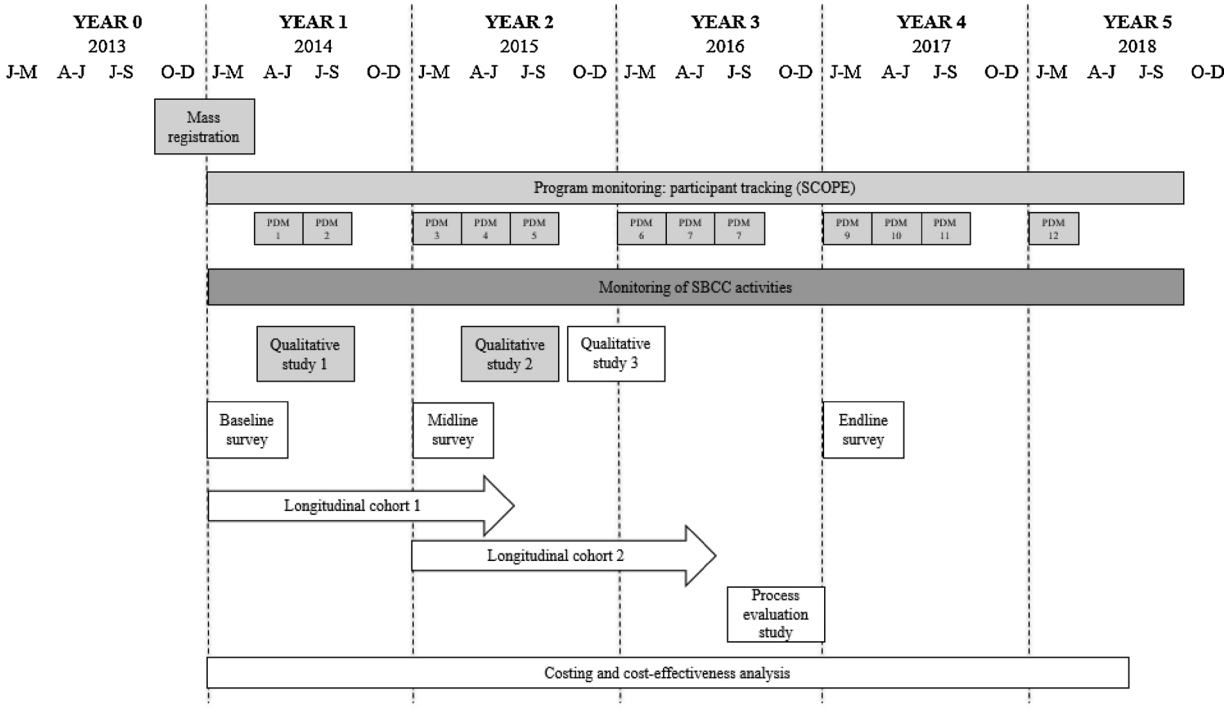

Fig. 3. Timeline and responsibilities of monitoring and evaluation activities for the RFRT program in Ntchisi District, Malawi.

monthly, annually, and over the course of the program. For example, the mean enrollment over the four years of implementation was 82%, with 63% in year 1 (2014), to 79.3% in year 2 (2015), 89% in year 3 (2016), and 90.5% in year 4 (2017). Data visualization was accomplished through simple graphs and tables, which could be accessed simultaneously by program planners and managers at the local, country, regional and headquarter WFP offices, who collaborated in real-time to make evidence-informed program implementation decisions. Some of these decisions included, for example: (1) the addition of distribution sites in areas that exhibited low attendance; (2) directed, individual-level follow-up with households that demonstrated repeated absences; (3) improved age-based beneficiary targeting (6–23 months); and (4) more efficient tracking and purchasing of SQ-LNS that led to program cost savings.

3.2. Post-distribution-monitoring (PDM) surveys

PDM surveys were led and conducted every four months by WFP, in January, May, and September, beginning in 2014. They comprised a second and complementary monitoring mechanism to SCOPE. PDM surveys were used to monitor output and outcome indicators throughout the year, which generated repeated measures on a number of important indicators that were compared across seasons as well as the project period. Specifically, PDMs collected data on self-reported SQ-LNS adherence and use (consumption, sharing, selling, theft), as well as knowledge of WASH and IYCF messages and practices. The purpose of repeat measurements of these intermediate outcomes was to capture whether and how practices varied over time and by seasons. Monitoring of actual consumption of the food product was not measured. IYCF indicators minimum meal frequency (MMF), minimum dietary diversity (MDD), and minimum acceptable diet (MAD) were assessed in the surveys from January 2015 and the household food security indicator Food Consumption Score Nutrient Quality Analysis (FCS-N) from January 2016 to respond to evolving program interventions, such as the addition of the small livestock and home gardening component to increase dietary diversity of program beneficiaries. PDM surveys used a two-stage stratified sampling method to identify a new, representative sample of 350 mothers of children aged 6–23 months and 350 pregnant women at each survey round. The sampling frame used for each PDM survey was based on district population data and program registration lists. In the first stage of sampling, clusters (group village heads [GVH], a traditional aggregation of villages in Malawi) were randomly selected from district population data using probability proportional to size sampling. In the second stage, mothers of children 6–23 months and pregnant women were randomly selected for inclusion in the survey. PDM sample sizes were calculated based on a 65% estimated prevalence of program participation and relative precision of 0.065 for the participation prevalence estimate.

3.3. Training for PDM surveys

Training and pre-test data collection for the surveys was held three times per year, prior to data collection. National enumerators were recruited and selected based on their understanding and knowledge of the local context. To foster familiarity and comfort with the interviewees, all interviews were conducted in the local language, Chichewa, using mobile devices (tablets) that promoted rapid data compilation and analysis. During the two-day training, indicators included in the survey, interviewing techniques, and food group names in the local language were reviewed and studied. The training also allocated time for familiarization of the mobile survey device, discussion sessions, and a half day of pre-tests in the field. Approximately 90% of the enumerators who participated in a PDM survey had participated in other WFP-led local surveys, making them familiar to the context and general interviewing technique.

Complete PDM reports were disseminated within two weeks of data collection. The reports analysis, in complement with the SCOPE dashboard results, were consistent with the planned efforts to implement near real-time, evidence-based course correction and adaptive programming.

Data collected through PDM surveys was also used for various analytical purposes (publications forthcoming), including (1) the

Table 1
Objectives of the JHU-led impact evaluation of the Right Foods at the Right Time program.

Research Objective	Methodology used	Study population
1. To assess the difference in the prevalence of stunting (% below -2 z-scores in length for age, LAZ) among children 6-23 months of age in the program versus comparison areas	Repeat cross-sectional surveys in program and comparison districts	Children 6-23 months
2. To assess the difference in the 18-month pattern and rate of linear growth, from 6-23 months of age, in a longitudinal cohort of children in the program versus comparison areas	Longitudinal follow-up of children 6-23 months, beginning at 6 months of age and every 6 months (at 12, 18, and 23 months) until 23 months of age in the program and comparison districts	Children 6-7.5 months at baseline (2014) and midline (2015) cross-sectional surveys
3. To assess the difference in the prevalence of anemia in children and pregnant and lactating women (PLW) in the program versus comparison areas	Hemoglobin measurements in the program and comparison districts at baseline (2014) and endline (2017)	A subsample of PLW with children < 6 months, and children 6-23 months
4. To assess the changes in maternal knowledge, attitudes and practices related to IYCF and PLW in the program versus comparison areas	Repeat cross-sectional surveys in program and comparison districts	Children 6-23 months
5. To identify facilitators and barriers to program access, participation, and uptake, and community perceptions of the program and its benefits	Qualitative study that included: In-depth interviews (IDIs) Free listing of foods Pile sorting of foods Direct observations Focus group discussions (FGDs)	Mothers and caretakers (fathers, grandmothers) of children 6-23 months Village leaders Program staff, including nutrition Promoters and Care Group Volunteers (CGV)
6. To assess the cost effectiveness of the program focusing on outcomes of stunting and disability-adjusted life years (DALY)	Cost-effectiveness study to measure the incremental cost effectiveness ratio (ICER) to assess the cost per case of stunting averted	

relationship between dietary diversity and market accessibility; (2) the relationship between diarrhea prevalence and WASH practices; (3) the sensitivity of IYCF indicators of MDD and MAD to seasonal differences and how these indicators could be used in surveillance systems for early warning; (4) the exploration of associations between household food security, measured by the FCS-N, and individual-level nutrition indicators MAD and MDD; and (5) the process evaluation study (described in more detail below).

3.4. Monitoring of input and output indicators by the cooperating partners

WVM, who were the cooperating partner for implementation of the program, provided regular monitoring on the RFRT through monthly reports on input and output indicators, such as the completion of training by program implementing staff.

3.5. Qualitative studies

Between 2014 and 2017, three qualitative studies were conducted with the objective of informing programmatic course correction (Fig. 3). WFP led the first two qualitative studies, from May-July 2014 and 2015, respectively, to explain gaps identified in PDMs as well as from WFP and WVM staff that warranted further investigation. More specifically, the first study assessed the key drivers of program implementation and coverage of the SBCC component, with a focus on front-line implementers (care group volunteers) for decision-making and action. The second study sought to understand the reasons for discordant results of high IYCF knowledge but suboptimal IYCF practices identified in PDM surveys. The results of the second qualitative study, in combination with PDM survey results, were the impetus behind the inclusion of the nutrition-sensitive program component that was added in year two of the program's implementation.

The third qualitative study, led by JHSPH, was conducted from November to December 2015 and sought to understand the facilitators and barriers faced by mothers of eligible program participants (children 6–23 months of age) as it related to attendance and participation at monthly SQ-LNS distributions, as well as the household-level implementation of SBCC messages around use of SQ-LNS, and IYCF and WASH knowledge, practices, and behaviors (Ruel-Bergeron et al., 2018).

All three qualitative studies used a combination of in-depth interviews (IDI), and focus groups discussions (FGD), which were the primary methods of data collection, given the flexibility with which they convey depth and understanding of certain issues of interest. The JHSPH-led qualitative study also included free listing, pile sorts, and direct observations, as well as IDI and FGD with other households (fathers and grandmothers), community members (village leaders), and program staff (nutrition Promoters and Care Group Volunteers). The inclusion of household and community members broadened our understanding of the context in which mothers made decisions about their child's nutrition and care, while the inclusion of program implementing staff enabled a better understanding of their responsibilities in program delivery and implementation. The combination of various qualitative research methods and respondents was used to provide context, as well as to complement and triangulate information collected during the IDIs on a range of maternal and child knowledge and behaviors. Sampling, analysis, and results of these qualitative studies are forthcoming publication (Ruel-Bergeron et al., 2018).

4. Impact evaluation of the Right Foods at the Right Time

The independent impact evaluation of the RFRT was led by JHSPH and aimed to measure the impact of the program in improving linear growth and reducing stunting among children 6–23 months of age in the program district after three years of implementation. Given the known complex and distal nature of stunting to individual-level feeding behaviors, as well as arguments against the use of stunting as an individual-level classifier of undernutrition (Perumal, Bassani, & Roth, 2018), the impact evaluation also included more proximal measures of nutritional status and behaviors, such as child anemia, underweight, wasting, and infant and young child feeding practices that might be more responsive to the interventions delivered by the program (objectives 3 and 4, Table 1).

The six research objectives of the evaluation and respective research activities used to respond to them are outlined in Table 1. Given the GoM's targeted and purposeful selection of the program district, a comparative design was used. Repeat cross-sectional surveys administered to pregnant and lactating women (PLW) and program-eligible children responded to objectives 1, 3, and 4; a longitudinal cohort of children 6–23 months of age was enrolled to respond to objective 2; a qualitative sub-study responded to objective 5 (described above); the compilation of costing data was used to respond to objective 6; and a process evaluation was used to assess timely recruitment into the program, program fidelity (including quality), and reach of the program's

Table 2
Outcome measures for the impact evaluation study, as collected in cross-sectional surveys and longitudinal cohorts.

Outcome	Instrument	Methodology
Child and PLW nutrition outcomes		
Length	Seca 417 length boards (child) Seca 213 stadiometer (PLW)	Supine length (child) or standing height (PLW) measured to the nearest 0.1cm 3 repeated measurements per child/PLW
Weight	Seca 374 digital infant scale (child) Seca 876 flat scale (PLW)	Weight measured to the nearest 0.1 kg One measurement per child/PLW
Head circumference	Flexible plastic tape measure	Measured 3 times to the nearest 0.1cm
Mid-upper arm circumference	MUAC tape measure	Measured 3 times to the nearest 0.1cm
Anemia/ Hemoglobin	Hb301 hemoglobinometer (HemoCue AB, Angelholm, Sweden)	1 spot of whole capillary blood, collected from heel (child) or finger (PLW)
Infant and young child feeding outcomes[a]		
Dietary diversity score (DDS)	24-hour recall of dietary intake (food frequency) on 9 liquid (beverage) items and 16 food groupings	Calculation of a food group score (range 0-7) based on the consumption of foods falling into the following food groups: (1) grains, roots, and tubers; (2) legumes and nuts; (3) dairy products (milk, yogurt, cheese); (4) flesh foods (meat, fish, poultry and liver/organ meats); (5) eggs; (6) vitamin A-rich fruits/vegetables; (7) other fruits and vegetables. 1 point is assigned for each food group consumed.
Minimum dietary diversity (MDD)		Defined as the consumption of 4 or more food groups in the previous 24 hours (based on DDS ≥ 4)
Minimum meal frequency (MMF)		Defined as the consumption of 2 meals among breastfed children 6-8 months; consumption of 3 meals among breastfed children 9-23 months; consumption of 4 meals among non-breastfed children 6-23 months.
Minimum acceptable diet (MAD)		Defined as children who achieve both MDD and MMF.
Child morbidity		
Diarrhea, fever (high and low), symptoms of acute respiratory infection, malaria, eye infection, ear infection	Maternal report of child symptoms in the 30 days preceding the interview, and 7 days preceding the interview, among those who experienced symptoms in the last 30 days	Measured as 30-day and 7-day morbidity for each of the symptoms included

[a] All IYCF indicators are based on the 2008 WHO methodology (The World Health Organization (WHO) et al. (2008)).

implementation. This impact evaluation design comprehensively documented and measured various outcomes of interest (Table 2), and increased the certainty with which changes in child nutrition and maternal knowledge and practices can be attributed to the program's implementation.

4.1. Cross-sectional household surveys

The independent program impact evaluation included three cross-sectional surveys, all of which were conducted in the program and a neighboring comparison district, Dowa, from January to March 2014 (baseline), 2015 (midline), and 2017 (endline) (Fig. 3). The comparison district was selected given its topographical, climactic, demographic, and agricultural similarities to the program district. Household, child, and maternal questionnaires were administered to households with children 6–23 months and PLW, respectively.

The sampling frame for the program district consisted of all villages with a total population of 150–600 in 12 health catchment areas in the district, and 14 out of the 23 health catchment areas in the comparison district. In the comparison district, 9 out of 23 health center catchment areas were purposively excluded from the sampling frame for the following reasons: (1) to minimize potential contamination from the program district; (2) topographical differences from the program district and proximity to the capital city of Malawi (Lilongwe); and (3) home to a large refugee population.

Sample size calculations for the cross-sectional surveys were based on a conservative value of 50% given that the most recently published prevalence of stunting in Malawi at the time was 47% (National Statistical Office, ICF Macro, 2018), and a desired detection of an absolute difference in prevalence of 5%–7% stunting between the program and comparison districts at endline. The level of significance (alpha, Type I error) was set at 0.05, the power (1- type II error) to detect the difference in the stunting prevalence was assumed at 80%, and 10% loss to follow-up was accounted for. The required sample size was estimated at 1200 children aged 6–23 months per group, or 2400 per survey, and 7200 across the three surveys. PLW were estimated to account for 8–9% of the total population in the district. Thus, the number of villages selected to yield the required number of children were predicted to yield ≈3600 PLW per survey, of which only 1/3 (n = 1200) would be systematically sampled. The sample size for bio-specimen collection (e.g. hemoglobin) was set at 150 children and 150 PLW per district and for the baseline and endline surveys only, yielding n = 600 of each. Subsamples for bio-specimen collection were selected using a systematic sampling procedure of every eighth child and every fourth PLW selected.

All households in sampled villages were visited and asked to describe their household composition; those with children 6–23 months of age and/or PLW (< 6 months post-partum) were confirmed as eligible by verification of the child's date of birth as written in their health passport. All households with children 6–23 months of age and one in every two to three PLW were then revisited and administered three questionnaires: (1) a household composition module with name, age, and sex of all household members; (2) a household socioeconomic module with basic demographic, economic, and asset ownership information; and (3) a child or PLW module with anthropometrics, diet, morbidity, exposure to health and nutrition services, and consumption and use of SQ-LNS (child only), including whether any of the supplement was sold, shared, stolen, or lost. Though not all surveys reached their targeted sample size (n = 1200 children 6–23 months per district and n = 600 PLW per district), all surveys enrolled 93% or more of the target sample size. Flow diagrams for each of the cross-sectional studies, with the number of villages and children sampled for each district will be published in subsequent evaluation papers.

Anthropometric measurements were performed by trained, standardized anthropometrists using standard equipment and procedures. Standardization of all anthropometric measurements was established by having each measurer take repeated (2) measurements of ten children to calculate inter- and intra-rater reliability. The five highest-

performing measurers were responsible for survey measurements. Middle-upper arm circumference (MUAC), head circumference, child length, and PLW height were obtained and recorded in replicates of three. Weight was measured one time using a Seca 374 digital infant scale for children and Seca 876 flat scale for PLW. Length was measured 3 times using Seca 417 length and Seca 213 height boards for children and PLW, respectively. Hemoglobin was assessed using HemoCue model 301 machines (HemoCue AB, Angelholm, Sweden), by trained phlebotomists certified in HIV-voluntary counseling and training (VCT). Immediate results of the hemoglobin test were recorded on the participant's questionnaire form, and participants identified as being anemic (Hb < 11.0 g/dl among children and lactating women, and Hb < 10.0 g/dl among pregnant women) were referred to the nearest health center.

4.2. Longitudinal cohorts of children 6–23 months

All children who were 6–7.5 months of age during the baseline or midline surveys were enrolled into a longitudinal cohort (Fig. 3). The inclusion of longitudinal cohorts in the study design was to enable the tracking of individual children for the entire period of program eligibility in the program district, and to examine how and whether their linear growth trajectories differed from those of children in the comparison district over time. Children in the longitudinal cohorts were followed in six-month intervals from 6 to 24 months of age (at 6, 12, 18, and 24 months), for anthropometric, diet, morbidity, and program use and participation assessments (Fig. 3). Tracking and identification of children enrolled in the longitudinal cohorts was done through a combination of methods, including household global positioning system (GPS) coordinates and cross-verification of household contact information, hand-written maps, and study ID number as marked on the household's exterior and inside of the child's health booklet.

Based on the cross-sectional sample size of 1200 children per district, and an estimated 5.5% of those children being six months of age at the time of the survey, the sample size for the longitudinal cohorts were expected to be 66 children per district, per cohort (n = 132 children enrolled at baseline, and n = 132 enrolled at midline), yielding a total of 264, six-month-old children with longitudinal growth data. Since the longitudinal study's sample size was based on the cross-sectional sample size, it was not highly powered to detect changes in linear growth.

4.3. Process evaluation

The process evaluation combined data from multiple sources to construct an assessment of the program's operations, implementation, and uptake. Three major categories of process data were included (Table 3), and were based on Linnan and Steckler's definitions (Linnan & Steckler, 2002), as follows: (1) timely program recruitment, which compared actual time of program uptake by eligible beneficiaries to the optimal; (2) program fidelity, which encompassed aspects of quality of delivery, dose delivered, and dose received; and (3) program reach, or coverage. The SCOPE and PDM monitoring systems provided program recruitment, fidelity, and reach/coverage data. These data were complemented by structured direct observations of SQ-LNS distributions (n = 14) and one-on-one counseling sessions by program implementers (n = 14), which assessed program quality, such as functioning of the participant tracking monitoring systems, and communication of IYCF messages by program staff. Additionally, structured questionnaires were administered to front line program staff (n = 14 Promoters and n = 70 Care Group Volunteers) to measure their knowledge of key IYCF, WASH, and SQ-LNS messaging.

4.4. Costing and cost effectiveness

The costing analysis had two primary objectives. Firstly, to measure the costs associated with the implementation of the nutrition program in the program district. As the Government of Malawi continues to work towards scaling up nutrition activities in other districts across the country, knowing the costs associated with the implementation of such a program was deemed as critical, not only for budgeting purposes, but for leveraging of additional resources and support. The second objective of the costing work was to measure the cost-effectiveness/consequences of the intervention vis-a-vis the primary program outcomes and impact. These included, for example, cost effectiveness of improving diet, reducing stunting, and promoting linear growth.

4.5. Training and quality control for impact evaluation activities

All research assistants were locally hired, spoke the local language, Chichewa, and participated in a five-day training preceding each cross-sectional survey, with one additional day for an anthropometry standardization exercise. Longitudinal cohort follow-ups were conducted by a sub-group of seasoned research assistants who had participated in the cross-sectional surveys, and who performed anthropometric measurements within an acceptable margin of error (± 5 mm for length and height, and ± 0.1 kg for weight). Training for the JHU-led qualitative study was also five days duration, and for the process evaluation study, two days in duration.

Quality control was maintained in all impact evaluation-related research activities by a program coordinator and locally hired senior-level study supervisor, both of whom accompanied study teams in the field, and who performed systematic quality control checks on a subsample (> 75%) of completed questionnaires, interviews, and/or transcripts. Weekly or daily debriefings with research assistants and throughout data collection were conducted to discuss unique situations that arose in the field, and concerns or queries regarding the wording of questions, and coding or completion of questionnaire modules. Quality of interviewing and transcription skills for qualitative activities was ensured by review of recorded interviews and corresponding translation and transcription by a supervising staff member fluent in the local language. Digital scales were checked daily for accuracy using standard weights (5 and 9 kg); recalibration was performed for scales that provided readings that were +/- 0.3 kg. Quality control checks were done weekly for stadiometers using a 150 cm standard measuring tape; stadiometers that were off by +/- 0.2 were replaced. Observations of

Table 3
Definitions used for process evaluation indicators.

Process evaluation component	Definition
Recruitment	Procedures used to approach and attract participants. Recruitment often occurs at the individual and organizational/community levels.
Fidelity	The extent to which the intervention was delivered as planned. It includes three critical elements of quality, dose delivered, and dose received.
Quality	The quality and integrity of the intervention as conceived by the developers.
Dose delivered	The number of intended units of each intervention or each component that are delivered
Dose received	The extent to which the target audience (of the dose delivered indicator) actively engages with, interacts with, is receptive to, and/or uses materials or recommended resources.
Reach	The proportion of intended target audience that participates in an intervention, often measured by attendance. Reach is a characteristic of the target audience.

anthropometric and hemoglobin testing procedures were done periodically by supervisors using a quality control checklist to ensure quality among staff conducting height, weight, and hemoglobin measurements.

4.6. Ethical procedures and approval of the impact evaluation study

Informed consent was administered in the local language, Chichewa, and obtained from all study participants included in the various study components described above. Consent was obtained by fingerprint on the consent form among illiterate respondents.

Research assistants trained in ethical matters collected all data. They maintained the confidentiality of participants by keeping completed questionnaires stored away during field work, and submitting completed questionnaires to the team supervisors at the end of each day. All data at the analysis and reporting stage were de-identified by removing all household member first and last names, district name, health center name, and village name. Unique ID numbers that concatenated district, village, household, and individual pre-assigned codes were generated for each individual included in the dataset. The link between the unique ID and individuals and was not known by data analysts or researchers. The link was stored electronically and only senior investigators had access. All standard data privacy and protection mechanisms were used to ensure fidelity of the confidentiality of participants.

The impact evaluation study was approved by institutional review boards at JHSPH (IRB #00005237) and the College of Medicine Research and Ethics Committee (COMREC) (IRB #P.09/13/1469) in Malawi. The impact evaluation is also registered at ClinicalTrials.gov with #NCT 02985359.

5. Conclusion

This paper described the monitoring and evaluation that was applied to assess the implementation and impact of the stunting prevention program in Ntchisi District, Malawi. The use of mixed methods and M&E systems allowed the GoM, program implementers and evaluators to comprehensively understand the context in which community members and caretakers make decisions, the program's quality of implementation, and why and how impact could be achieved or not with the chosen intervention components in this setting. The documentation of program implementation and impact is seen as being especially important to isolate the impact of each program component in this context given the limited impact of SQ-LNS (or similar products) and nutrition education on child growth in Malawi to date (Kuchenbecker, Reinbott, Mtimuni, Krawinkel, & Jordan, 2017; Phuka, Maleta, & Thakwalakwa, 2008; Phuka, Maleta, & Thakwalakwa, 2009). For instance, group-level, village-based nutrition education to caregivers on key IYCF and WASH practices demonstrated impact on intermediate outcomes of child dietary diversity but not stunting in a recent study conducted in Northern Malawi (Kuchenbecker et al., 2017). The results of the impact evaluation, in combination with complementary program monitoring, contribute to an improved understanding of whether and how the package of interventions selected for this program was or was not effective in this context, which aspects of growth were affected based on the various measures collected, and which considerations should be taken into account in future program scale up and expansion. Further, the inclusion of various and complementary study activities strengthens the internal and external validity of both the evaluation's findings and the program's implementation.

A limitation to this program's impact evaluation is the lack of a randomized design. However, such a design was not only deemed inappropriate for documenting program effects in practice, it was also not possible given the GoM's targeted selection of a program district. The use of multiple and concurrent monitoring and evaluation activities, however, contributed to the evaluation's rigor. The impact evaluation design also presented valuable design and analytic alternatives to randomized designs that are often unacceptable and infeasible in the evaluation of large government-implemented programs as this one. As a nutrition community and in our efforts to scale up nutrition efforts globally, the use of innovative monitoring and evaluation systems and designs that rigorously document the effectiveness of a package of interventions generates important lessons for researchers and program planners alike.

5.1. Lessons learned

A number of lessons emerged as both program planners and evaluators reflected on the design and implementation of this comprehensive M&E system:

1 The absence of inclusion of an older cohort of children (24–41 months) in the impact evaluation limited our ability to measure program impact on children who had received full exposure (18 months, from 6 to 24 months of age) to the RFRT. As described by Menon et al. (2013), if the main period of growth faltering occurs before 24 months, as in Malawi, the greatest impact of the program on child height would have been observed once they have completed their full exposure to the intervention, i.e. > 24 months (Menon, Rawat, & Ruel, 2013). Thus, in a future study evaluating a similar intervention, the addition of an older cohort on which to measure anthropometric and biochemical outcomes would be important. Alternatively, the use of a longitudinal, as opposed to a cross-sectional design, though difficult in this setting, would more effectively demonstrate differences in child growth between intervention and comparison groups.
2 The ongoing training, motivation, supervision, and encouragement of care group volunteers is critical to program success and a key element for high retention of beneficiaries throughout the life of the program. Given the volunteer nature of the position, these individuals often have many competing priorities, so emphasizing their key role in the overall system and providing continuous support to enable them to do their job is essential to successful implementation and program impact.
3 The experiences from trainers and field supervisors was that seasoned research assistants were more capable and efficient in collecting data in the field. This observation led us to the conclusion that minimizing turnover of research assistants and/or ensuring the presence of the same research assistants, when possible, facilitated the acquisition of skills among new staff and reduced the length required for mastery, standardization, implementation of data collection protocols and procedures.
4 Investing in understanding the country and program context, including the behavioral patterns of the population is necessary for implementing the most efficient M&E systems and reaching the target population. In Malawi, seasonality is an important driver of population mobility, and understanding this a priori could have facilitated the identification of certain population groups (PLW, for example).
5 The experiences of the implementation of the M&E systems in this program reiterate the known importance of having complementary monitoring and evaluation systems (Levinson et al., 1999). Consistency, collaboration, and alignment among partners implementing various components of comprehensive M&E systems allowed triangulation of data, increased knowledge sharing, supported prompt course correction for program improvements, and framed evaluation results in the appropriate programming context.
6 The careful timing of monitoring and evaluation activities is important and provided a constant flow of information and feedback into the program. The use of these timely results, especially from monitoring systems, allowed for multiple course corrections and adaptive programming that was critical to improving access and

acceptance of the program's various components.

7 The use of developments in technology, such as tablets, computer, and mobile phone-assisted monitoring systems greatly facilitated the functionality of monitoring activities and provided near real-time data visualization that contributed to program adaptations. Combining these data sources with WVM output indicators onto one platform allowed program implementers to remain informed of all progressions of the program, and greatly contributed to rapid and streamlined data collection and analyses. These technology-based solutions also automated data management and calculation of indicators, reducing the risk of human error.

Funding source

This work was funded by the Children's Investment Fund Foundation (CIFF), London, UK. All program and study design decisions described in this manuscript, as well as decision to submit the article for publication were agreed upon with the funding source.

References

Black, R. E., Allen, L. H., Bhutta, Z., et al. (2008). Maternal and child undernutrition: Global and regional exposures and health consequences. *Lancet, 371*(9608), 243–260.

Black, R. E., Victora, C. G., Walker, S. P., et al. (2013). Maternal and child undernutrition and overweight in low-income and middle-income countries. *Lancet, 382*(9890), 427–451.

Chikhungu, L. C., & Madise, N. J. (2014). Seasonal variation of child under nutrition in Malawi: Is seasonal food availability an important factor? Findings from a national level cross-sectional study. *BMC Public Health, 14*, 1146.

Espo, M., Kulmala, T., Maleta, K., Cullinan, T., M-LL, S., & Ashorn, P. (2002). Determinants of linear growth and predictors of severe stunting during infancy in rural Malawi. *Acta Paediatrica, 91*(12), 1364–1370.

Kuchenbecker, J., Reinbott, A., Mtimuni, B., Krawinkel, M. B., & Jordan, I. (2017). Nutrition education improves dietary diversity of children 6-23 months at community-level: Results from a cluster randomized controlled trial in Malawi. *PloS One, 12*(4), 1–19.

Levinson, F. J., Rogers, B. L., Hicks, K. M., Schaetzel, T., Troy, L., & Young, C. (1999). Monitoring and evaluation of nutrition programs in developing countries. *Nutrition Reviews, 57*(5), 157–164.

Linnan, L., & Steckler, A. (2002). *Process evaluation for public health interventions and research* (1st ed.). San Francisco, CA: Jossey-Bass.

Menon, P., Rawat, R., & Ruel, M. (2013). Bringing rigor to evaluations of large-scale programs to improve infant and young child feeding and nutrition: The evaluation designs for the Alive & Thrive initiative. *Food and Nutrition Bulletin, 34*(3 Suppl), S195–211.

National Statistical Office, ICF Macro (2017). *Malawi demographic and health survey 2015-16.* Malawi and Rockville, MD: Zomba.

National Statistical Office, ICF Macro (2011). *Malawi demographic and health survey 2010.* Zomba, Malawi, and Calverton, Maryland: NSO and ICF Macro.

Olofin, I., McDonald, C. M., Ezzati, M., et al. (2013). Associations of suboptimal growth with all-cause and cause-specific mortality in children under five years: A pooled analysis of ten prospective studies. *PloS One, 8*(5).

Pensulo Phiri, F. (2016). *Experiences of multi-sector programming in Malawi.* F Exch June (52).

Perumal, N., Bassani, D. G., & Roth, D. E. (2018). Use and misuse of stunting as a measure of child health. *The Journal of Nutrition, 148*(3), 311–315.

Phuka, J. C., Maleta, K., Thakwalakwa, C., et al. (2008). Complementary feeding with fortified spread and incidence of severe stunting in 6- to 18-month-old rural Malawians. *Archives of Pediatrics & Adolescent Medicine, 162*(7), 619–626.

Phuka, J. C., Maleta, K., Thakwalakwa, C., et al. (2009). Postintervention growth of Malawian children who received 12-mo dietary complementation with a lipid-based nutrient supplement or maize-soy flour. *The American Journal of Clinical Nutrition, 89*, 382–390.

Prendergast, A. J., & Humphrey, J. H. (2014). The stunting syndrome in developing countries. *Paediatrics and International Child Health, 34*(4), 250–265.

Ruel, M. T., & Alderman, H. (2013). Nutrition-sensitive interventions and programmes: how can they help to accelerate progress in improving maternal and child nutrition? *Lancet, 382*(9891), 536–551.

Ruel-Bergeron, J. C., Hurley, K. M., Kapadia-Kundu, N., et al. (2018). Physical and sociocultural facilitators and barriers to access and utilization of a nutrition program in rural Malawi: a qualitative study. *Ecology of Food and Nutrition, 57*(5), 405–424.

Scaling Up Nutrition (2011). *Scaling up nutrition [Internet]. Washington, D.C.* [cited 2015 Jun 9] Available from:http://scalingupnutrition.org/sun-countries/malawi#tab-3.

The World Bank (2017). *Malawi [Internet].* [cited 2017 Feb 8], Available from:http://www.worldbank.org/en/country/malawi.

The World Food Programme (WFP) (2014). *SCOPE in five minutes.* Rome, ItalyWFP.

The World Food Programme (WFP) (2015). *How WFP supported the scaling up nutrition (SUN) roll-out in Malawi: A case study on stunting prevention programming at scale.* Rome, Italy.

The World Health Organization (WHO), The United Nations Children's Fund (UNICEF), The United States Agency for International Development (USAID), The Academy for Educational Development (AED), The Food and Nutrition Technical Assistance (FANTA), The International Food Policy Research Institute (IFPRI) Indicators for assessing infant and young child feeding practices. Part I: Definitions Geneva, Switzerland 2008.

United Nations Children's Fund (UNICEF) (1990). *A UNICEF policy review: Strategy for improved nutrition of children and women in developing countries.* New York, NY.

United Nations Development Programme (UNDP) (2016). *Human development report 2016*New York, NY.