Assessing the impact on child nutrition of the Ethiopia Community-based Nutrition Program

Report to UNICEF of an evaluation study, September 2012

Jessica White, MPH, and John Mason, PhD

Tulane University School of Public Health and Tropical Medicine,
New Orleans

TABLE OF CONTENTS

TABLES AND FIGURES

ABBREVIATIONS

ANC	Antenatal Care
CBN	Community Based Nutrition
CWS	Community weighing session
DHS	Demographic and Health Survey
EA	Enumeration Area
EDHS	Ethiopian Demographic and Health Survey
EHNRI	Ethiopian Health and Nutrition Research Institute
ENA	Essential Nutrition Actions
ENCU	Emergency Nutrition Coordination Unit
EOS	Enhanced Outreach Strategy
FHC	Family Health Card
FMOH	Federal Ministry of Health
GoE	Government of Ethiopia
HAZ	Height-for-age z-scores
HEP	Health Extension Program
HEW	Health Extension Worker
HF	Health facility
HHS	Household Hunger Scale
IYCF	Infant and Young Child Feeding
NCHS	National Center for Health Statistics
NNP	National Nutrition Program
NNS	National Nutrition Strategy
OHC	Other Health Card
ORS	Oral rehydration salts
PAD	Project Appraisal Document
PIM	Project Implementation Manual
PLW	Pregnant and lactating women
ppts	Percentage points
ppts/yr	Percentage points per year
PSNP	Productive Safety Net Program
RCT	Randomized control trial
SES	Socioeconomic status
TSF	Targeted Supplementary Feeding
UNICEF	United Nations Children's Fund
VCHW	Volunteer Community Health Worker
WAZ	Weight-for-age z-scores
WFP	World Food Program
WHO	World Health Organization
WHZ	Weight-for-height z-scores

ACKNOWLEDGMENTS

This evaluation study brings together the work of many people. First, the people who designed and implemented the program should be recognized for having started and expanded a program that reached individual mothers and children across a large part of Ethiopia – probably a more widely implemented nutrition program than anywhere in Africa. The Government of Ethiopia, Federal Ministry of Health, with the Ethiopia Health and Nutrition Research Institute (EHNRI), with extensive training and support by UNICEF in the field, launched the National Nutrition Program in 2008, with financial assistance from a World Bank loan.

Second, four surveys were carried out, by Addis Continental Institute of Public Health, and Mela Research PLC. These surveys, covering 60-61 different enumeration areas each, involved extensive logistics and persistence, often in difficult country and remote villages. Their diligent work is recognized as the basis for this evaluation research. The preliminary reports from the baseline and midline surveys[1] are the source of descriptive survey results.

Third, the weighing program data, although not a main focus of the research reported here, provides a continuing estimate of the extent of implementation and likely resulting trends in malnutrition. These data were compiled and analyzed at Tulane by Anne Hoblitt, Laura Buback, and Jessica White.

The evaluation work has benefitted from advice, consultation, and key background information from many people in UNICEF: at the beginning, Iqbal Kabir, Kyoko Okamura, and Bjorn Ljungqvist; more recently, Wigdan Madani, Joan Matji, Selamawit Negash, Adam Bailes, and Abebe Hailemariam.

From the World Bank, at the beginning Marito Garcia and Sunil Rajkumar; more recently Yuki Isogai, Matthew Robinson, Ferew Tekabe, and Zia Hyder

The team from the Tulane University School of Public Health and Tropical Medicine includes Profs Joe Keating and Kat Theall (epidemiology and evaluation design), Lisa Saldanha, Jessica White, and John Mason (PI).

The evaluation research has been supported financially by UNICEF and the World Bank.

Author Contact: Jessica White: jwhite8@tulane.edu ; John Mason: masonj@tulane.edu

[1] Addis Continental Institute of Public Health (2009 & 2010). Community-Based Sub-Component of Ethiopian National Nutrition Program: Baseline Survey Report; Mela Research PLC (2011). CBN Evaluation Survey: Ethiopia.

ABSTRACT

Background. This report is based on analysis of data collected under UNICEF supervision to assess the effectiveness of the Community-based Nutrition (CBN) component of the Ethiopian National Nutrition Program (NNP). The Government of Ethiopia (GoE) legislated a National Nutrition Policy in 2008. This centrally included a NNP, for which some $30 million were initially committed by the World Bank. The overall aim was to reduce malnutrition among mothers and children in one of the world's countries with highest malnutrition rates and subject to repeated droughts. The strategy for the first five years was to consolidate ongoing nutrition services, transitioning into sustainable interventions using the community-based approach. CBN programs were launched by area, groups of woredas (districts) starting each year from 2008, covering 228 woredas (out of about 800) by 2012. CBN activities are conducted by Volunteer Community Health Workers (VCHWs) in kebeles (villages), supervised by Health Extension Workers (HEWs), from health posts.

Objective. The research aimed to assess the plausible attribution of changes in anthropometry to CBN activities, and describe program implementation and selected process indicators. This is intended to be useful for (a) decisions on extending the program, a second phase having been planned for 2013-2018, and (b) to add to current knowledge of the effectiveness of such programs, for policy decisions on their utility in similar settings especially in Africa.

Methods. Following the launching of the CBN in tranches 2 and 3 in 2009-10, four evaluation sample surveys were carried out, two at the baselines and two at midline in September 2011. The four surveys covered about 60 randomly selected clusters each, re-sampling households from the same clusters (referred to throughout as EAs, as they were defined as census enumeration areas) at midline. Thus each EA acted as its own control, and the main outcome estimates were from changes within each EA. No comparison external groups were possible at baseline, according to the government's policy that no surveys could be done without programs immediately following. However *de facto* internal comparison groups were possible, from varying implementation. Changes found could also be compared to long-term trends. A number of potential confounders were estimated: economic (as assets), education, environmental, access to services, food insecurity ('hunger scale'), and others. Regression analysis (OLS) was used to control for the possible effects of these. Data were also available from routine program reporting from weighings; these are used here for guidance, but reported in detail elsewhere.

Results. The gap between surveys was 28 months for tranche 2 and 18 months for tranche 3; however not all woredas started immediately, and the actual period of implementation by baseline had medians of 21 and 6 months respectively. Moreover, for tranche 3, 28% of woredas had not started CBN activities by the midline survey, and these formed one internal comparison group. All tranche 2 woredas had been implementing CBN for at least 12 months by the time of the midline survey. Within implementing areas by midline, participation was estimated at around 30% (children weighed), but higher in terms of contacts with health workers (70-85%), use of family health cards (60%), and other indicators. Although 0-2 yr children were the target, many 2-3 year olds were weighed. Crucial to the potential impact is that the ratio of VCHWs to children in each tranche was in the 1:10 to 1:30 range, consistent with having impact.

Substantial changes in infant and young child feeding (IYCF) practices were observed. Exclusive breastfeeding under 6 months, already high, increased in tranche 2 to nearly 90%. Dietary diversity at 6-23 months increased significantly, as did the minimal acceptable diet at this age – these prevalences reaching around 40-50%. As further examples, the practices of providing less food during diarrhea, and eating less during pregnancy, were significantly reduced. Maternal use of antenatal care increased, and women taking iron-folate during pregnancy increased from 30 to 50% in tranche 2. Deworming during pregnancy however remained low, around 10%.

Stunting (and HAZ) were the most significant indicators associated with project implementation; underweight was similar but less robust, and wasting fluctuated uninterpretably (at lower prevalences). The long-term without-program stunting trend was of -1.3 ppts/yr. Compared with this, the stunting change rates in tranches 2 and 3 were significantly higher, at -4.3 and -5.3 ppts/yr – an additional 3-4 ppts/yr over the long-term trend (which was already improving). Differences in VCHW activity were used to assess program implementation in tranche 2, analogous to the non-implementation group in tranche 3, for comparisons.

Measures of SES (roofing), education, and toilet facilities all showed some improving changes over time; drinking water did not. These and other potential confounders were used in regressions, with differences in stunting (by EA) as the dependent variable; they were tested both as changes in these indicators (e.g. ppts change in good roofing), and as absolute values (at midline). The estimates of association (i.e. coefficients) of change in stunting with program implementation remained significant (p=0.04-0.07) and of similar size when controlling, providing evidence that the rates of improvement associated with the program activities were likely to be attributable causally to the program.

The Targeted Supplementary Feeding (TSF) program was assessed along with the CBN. Essentially, either or both programs produced the same improvement, and these were not additive. It was noticeable that both the TSF and the CBN were implemented in clusters with higher starting prevalences of stunting. This is assumed to result from the targeting of these programs. The without-program clusters (i.e. with no CBN or TSF) showed no significant improvement, in contrast to the with-program clusters.

The project plans put forward a number of quantitative aims that match indicators from the evaluation surveys. In all the process indicators of maternal and child participation, the aims were met or exceeded. The same applies to indicators of IYCF, except for the deworming during pregnancy. The aims in terms of stunting and underweight improvement were modest, and were greatly exceeded for stunting, and on target for underweight.

Conclusions. The CBN activities plausibly had a significant benefit on child anthropometry and other aims, well exceeding targets for example for stunting. The decisions that could be based on this are, for the Ethiopian NNP itself, to continue and expand activities, while correcting some weaknesses. The most important is the relatively low participation, around 30%, of children in woredas covered by the program. For policy decisions in Ethiopia and elsewhere, these results join and are supported by findings of other evaluations, gathering increasing evidence that community-based activities aimed at improving child nutrition can be effective. This experience in Ethiopia may be showing a way ahead for many communities in Africa and this needs policy support.

1. BACKGROUND

This report is based on analysis of data collected under UNICEF supervision intended to assess the effectiveness of the Community-based Nutrition component of the Ethiopian National Nutrition Program. The results will be more easily understood in the context of the program itself and the opportunities and constraints of evaluation, in both the Ethiopian and the global context. The report therefore begins with this background, in four sections:

A. A brief description of the Community-based Nutrition program (CBN), and aims of the evaluation
B. Design of the evaluation
C. A summary of the CBN routine data results
D. The global context of nutrition program experience and available evaluations.

A. The Community-Based Nutrition program, and evaluation objectives

The Government of Ethiopia (GoE) legislated a National Nutrition Policy in 2008. This centrally included a National Nutrition Program (NNP), for which by that time there was agreement in principle from the World Bank to provide an initial $30 million in funding over five years (July 2008 – June 20113), expected to be seed money for a much larger investment from other donors. The NNP plan – which had been prepared with World Bank and UNICEF input –laid out a program that was ambitious in terms of implementation (FMOH, 2008). The community-based component was linked to the much larger Health Extension Program (HEP), which trained some 30,000 new Health Extension Workers (HEWs) over the next few years. This was planned to cover most of the country's 600 woredas (districts, now increased to about 800) over five years. This in turn was a continuation and evolution of UNICEF and WFP's Enhanced Outreach Strategy (EOS) which provided a bi-annual campaign to deliver Vitamin A and deworming to children under-5 as well as screening for moderately malnourished pregnant and lactating women (PLW) and children under-5 for referral into the Targeted Supplementary Feeding program (TSF).

The objectives of the overall NPP of which CBN was one of four components, were stated as follows: 'The Project Development Objective (PDO) is to improve child and maternal care behavior, and to increase utilization of key micronutrients, in order to contribute to improving the nutritional status of vulnerable groups (World Bank, 2008, p 45)'. The roles of underweight and stunting were recognized as Higher Level Objective Indicators, but these 'are affected by several other factors beyond the control of the project, and target values have deliberately not been set… (p 29)'. However, under 'Key Performance Indicators' (p 90) aimed-for prevalences of underweight, stunting, and wasting are given, e.g. for stunting 46% at baseline and 40% at year 5 (a rate of improvement of 1.2 ppts/yr, which it turns out, as discussed below, is the same as the underlying without-program rate); we might take it that a with-program rate of 1.2 ppts/yr <u>above</u> the underlying rate is being suggested as an objective (i.e. 2.4 ppts/yr). Baseline, midline, and endline surveys were agreed to measure this (p 37).

The World Bank Project Appraisal Document (PAD) (quoted above) was approved in April 2008. At the same time the Government agreed on the NNP and a National Nutrition Strategy. The primary objective of the NNP was stated in the Project Implementation Manual (PIM) prepared by the Ethiopian Federal Ministry of Health (FMOH) as (FMOH, 2008 p 12-13):

'The Primary Impact Objective of the NNP is to improve nutritional and micronutrient status of the population especially mothers and children through cost effective and sustainable interventions. The achievement of the objectives is measured with the following impact indicators:

1. *Reduce the prevalence of underweight (W/A <-2) from 38% to 30% by 2013;*
2. *Reduce the prevalence of stunting (H/A <-2) from 46% to 40% by 2013;*
3. *Reduce the prevalence of wasting (W/H <-2) from 11% to 5% by 2013;*
4. *Reduce the prevalence of low birth weight (<2.5 kg) from 13.5% to 10% by 2013; or Reduce % of births considered smaller than normal from 28% to 24%;*
5. *Increase the proportion of pregnant women gaining at least 9 kg over the course of pregnancy.*

The Outcome Objectives include the following:

- *Increase the proportion of infants 0-6 months exclusively breast fed from 32 % to 60%;*
- *Increase the proportion of infants 6-9 months introduced to complementary food at 6-7 months from 25% to 50%;*
- *Increase the proportion of children with diarrhea who were fed "same or more than usual" from 25% to 50%;*
- *Reduce the prevalence of Bitot's spots in children aged 6 - 59 months from 1.7% to < 0.5%*
- *Reduce the prevalence of IDA in women of childbearing age from 26.6% to 15%.'*

These can provide more guidance on the impacts expected, presumed over the 5 year program planned. Primary Impact Objectives 1-3, and the first three under Outcome Objectives, can be readily assessed in surveys, and have been included here.

Concerning the strategy to reach the objectives described above (FMOH, 2008, p 13):

'The National Nutrition Program (NNP) seeks to address the above objectives using the following four main strategies:

1. *Consolidate and scale up ongoing national nutrition services;*
2. *Transition programs into preventative and sustainable interventions using the community-based nutrition approach;*
3. *Strengthen multisectoral nutrition linkages and;*
4. *Improve the capacity of institutions to formulate policies and implement the nutrition programme.*

On the basis of these strategies, the NNP I for the next five years (July 2008- June 2013) consists of two main components: a Nutrition Service Delivery component and an Institutional Strengthening component to support the service delivery (FMOH, 2008).'

The same quantitative targets are given both in the PIM (p 110-116) and PAD (p 90). Many of these have been assessed here and will be reported on later. In practice, the 'Primary Objective Indicators', which are anthropometric, and those for the 'Outcome Objective 1: Improve child and pregnancy feeding and caring behaviors,' are those most likely to be directly affected by the CBN component, and are taken as relevant to evaluating the CBN. These indicators are referred to individually later in the text.

The components of the NNP are given in table 1.1; in practice the first three were combined, as 'Nutrition Service Delivery', as foreseen in the strategy. Micronutrient interventions were separately (and less extensively) implemented. The initial plans for information and evaluation are covered by sub-component 2C.

The institutional arrangements of the CBN are as follows:

> *'The overall responsibility for CBN lies with the Woreda Health Office that will provide supportive supervision and technical support to implementation, with support from the respective Zonal and Regional Health Bureau. Under the Woreda Health Office, two HEWs per kebele will be supported by 10 -12 VCHWs each. VCHWs are community volunteers, preferably female, one per 50 households (FMOH, 2008, p 26).'*

The activities of the VCHW to improve nutrition include:

- Monthly weighing of children under 2 and counseling of mothers (triple A)
- Monthly community conversations (aimed at improving caring practices, hygiene, etc)
- Home visits to follow up growth faltering or sick children
- Referral of children sick or getting malnourished to Health Post, which can lead to therapeutic feeding
- Informal contacts within community (VCHW comes from and lives in the community).

Launching the program involved the regional, zonal, and woreda health offices; recruitment and training of VCHWs, provision of communication materials, scales, growth charts, and other program materials. UNICEF, in collaboration with the Health Bureaus at all levels, has taken a leading role in launching the program and in training and providing supplies. The program was launched each year by group of woredas, referred to as tranches, starting in 2008. The numbers of woredas involved and other details are given below in table 1.2, totaling to 228 in the first four tranches.

The aim of the evaluation research and analysis reported here is primarily to assess the effect attributable to CBN project activities, on outcome indicators, mainly anthropometric. In other words, the impact on young child nutritional status, measured by growth. Effects likely to be associated with the project, on infant and young child feeding (IYCF), maternal health, and certain process indicators have also been analyzed, but not (yet) further investigated in terms of their actual plausible causal attribution (this is for lack of time and resources for continuing analysis). These too are reported.

The estimated changes in indicators are discussed further in relation to the objective indicators in the plans (PAD and PIM), and to expected changes from other data from Ethiopia (e.g. DHS), and other countries.

B. Principles of the evaluation

The first condition for the evaluations planned was the policy of the GoE that surveys could only be carried out just before or during program implementation: no woredas without programs about to start or underway could be surveyed. Thus there could be no direct comparison groups for assessing with-without program effects. Even less, therefore, was randomization of intervention feasible.

The epidemiological literature has laid out carefully, first, the progression from efficacy results often based on randomized controlled trials (RCTs) of individual interventions, to using such results for large scale program planning involving multiple interventions through different platforms, to assessing impact by means other than RCTs in such programs, as randomization to intervention and control groups is rarely feasible in these (Habicht et al, 1999; Victora et al 2004). The efficacy of the interventions included – sometimes referred to as 'Essential Nutrition Actions'(USAID, 2011) – has been comprehensively reviewed as published in a series in the Lancet (Black et al, 2008; Victora et al, 2008; Bhutta et al, 2008; Bryce et al, 2008; Morris et al, 2008). All the activities included in the CBN have been scored in the Lancet series, which took data from 36 countries, as 'sufficient evidence for implementation in all 36 countries' (Bhutta et al, 2008, table 1). Thus the reasonable expectation is that if the interventions can be effectively implemented, there should be an impact on nutritional status.

Second, the ideas have been effectively laid out (Habicht et al, 1999; Victora et al 2004)and widely accepted in the health/nutrition field that traditional randomization to control and treatment groups in large scale operational programs is (a) not feasible and (b) that valuable information can be obtained by other methods. Experiments, such as RCTs, give 'probability' results; alternative methods (often described as quasi-experimental) applied when RCTs are impossible, give 'plausibility' results. In a plausibility study the case for attributable impact needs to be made, by comparisons between groups to which assignment is not randomized, taking account, statistically or otherwise, of possible threats to the plausibility.

This study fits exactly into the criteria of a plausibility study, as described by Habicht et al 1999, and Victora et al, 2004, and this concept is commonly seen in the literature[2].

The approach to evaluation put forward in 2008 (Mason, 2008: see figure 2.1 discussed in Methods) therefore sought compromises based on the constraints to randomization and budgetary restrictions, as well as the background data potentially available from Demographic and Health Surveys (DHS), from a National Nutrition Survey supported by the Bank and implemented by the Ethiopian Health and Nutrition Research Institute (EHNRI) in 2008. The basis of quasi-experimental design put forward by Cook and Campbell (1979) remains valid, and adaption of formal methods such as stepped wedge designs (Brown & Lilford, 2006) was applied. The design called for three evaluation surveys for each tranche (three were anticipated), at baseline, midline two years later, and endline five years after the baseline. A key feature was that each resurvey used the same clusters, and drew a new sample of children from each cluster: this allowed for each cluster to act as its own control.

2 E.g. 'Plausibility nutrition' gets 19,000 hits on Google Scholar. 'Plausibility health' gets 800 hits in PubMed.

Finally, evaluations are needed for decisions, preferably together with supporting data from other sources. In this case, two types of decisions were aimed for. First, the timing was aimed to give information for deciding on the extension, and possible modification, after five years for an expected second five year phase. These decisions are made in exercises like the mid-term and annual joint reviews of the project by the World Bank and the GoE. Initial results were made available in November 2011 (for the mid-term review), and those in this report in April 2012. Early results were communicated for adaptation of the program – for example on coverage from the CBN routine data.

Second, the results may contribute to decisions on carrying out such programs in *other* settings. Alone they will not be sufficient, but as part of a pattern from accumulating experience they could well contribute. The broader pattern of programs and evaluations is discussed in section ii of this background text.

C. CBN routine data results

The NNP began implementation of CBN in primarily agrarian woredas prone to food insecurity in four regions: Amhara, Oromia, SNNPR, and Tigray. The program rolled out in tranches that varied in timing and size, starting with 39 pilot woredas in 2008, a second tranche of 54 woredas in 2009, and a third tranche of 77 woredas in 2010 (see table 1.2 for details). In early 2010, routine results from the weighing programs, in villages and at health post started to flow to central administration. The extent was substantial and transmitted as Excel spreadsheets summarized at woreda level. Tulane University was asked to help with handling these extensive data, and by August 2010 the patterns that were appearing became clear, and were striking. They were also helpful in interpreting evaluation sample survey data. A summary graphic of tranche 1 data, by region and month from August 2008 to March 2012, is shown in figure 1.1: the original one of these, from 2008 -2010 from Hoblitt & Mason, 2010 was widely distributed. Two further reports updated these (Buback & Mason, 2011; White & Mason, 2011), and this process is continuing.

The results show a substantial drop in underweight prevalence soon after the regional programs start, of around 10 ppts, with a slower but still clearly improving trend thereafter. These prevalences apply to participants only. The coverage by woreda was high, and participation within woredas was estimated at about 30-40%. At population level this implied about 3-4 ppts initial drop. The ratio VCHWs to children under-2 ranged from 1:7 to 1:15 with an average of 1:11 (see table 1.3) (Buback & Mason, 2011; White & Mason, 2011). This is within the range expected to have an impact and is in line with the plans in the PIM, thus lending credence to the drop in prevalence observed in the CBN data.

These data came well after the evaluation surveys were designed so could not help the power calculations of sample size; but were not far different from the 5 ppts change the surveys were designed to detect. The shape of the curves indicated in figure 1.1 are also useful now for reviewing the plans for surveying new tranches, and for the endline surveys for the existing tranches (2 and 3).

D. Global context of programs and evaluations

Community-based programs are well known and considered likely to be effective in many poor settings as a platform for health and nutrition interventions. A recent review is in the Disease Control Priorities in Developing Countries, Volume 2 (Mason et al, 2006), and the key role of the VCHW in community health programs is discussed in Haines et al, 2007.

Within this, it has been suggested that the ratio of VCHWs to children is a key criterion, with 1:10-20 (part time VCHWs) considered ideal; this ratio depends on the part-timeness, or FTE, of the VCHW; this calculation is for 10% of time (FTE=0.1) (Haines et al, 2007). Recent work reviewing some 60 large scale health and nutrition programs worldwide, to be published by WHO as one of the World Health Assembly 2012 background papers (WHO, 2012), suggests a relation between impact (ppts/yr underweight prevalence reduction) and VCHW:child ratio, with 1:10-20 being well within the impact range, estimated as 2 ppts/yr improvement at 1:10 VCHWs:child. The graph is given in figure 1.2.

The WHO 2012 review of 60 large-scale programs identified 32 that had enough data to describe the program components. These showed that counseling and growth monitoring were common to most programs. Of these 32, eleven had enough data to estimate impact. (i.e. as in figure 1.2) The World Bank, in a volume entitled 'What can we learn from Nutrition Impact Evaluations?' (2010), assembled data on 28 programs, of which only six were community-based, of which two – Madagascar and Senegal – were actual large scale operational programs.

Programs similar to the CBN are well-established, although more so in Asia and Latin America than in Africa. Evaluations are less frequent and established.

2. METHODS

A. Design and data sources

An evaluation plan was developed for assessing overall changes during the five-year NNP I implementation period (Mason, 2008). The design, shown in figure 2.1, envisaged 9 surveys: base-, mid-, and end-line surveys, at intervals for three separate tranches of program roll-out. These followed a quasi-experimental evaluation design, aimed to estimate the impact on the population – more precise than, but usefully complemented by, the CBN routine program data. The design is a 'staggered implementation' approach, in which the trends within tranches of program implementation are compared with other trend estimates.

The evaluation was designed to estimate trends in malnutrition and in knowledge and practices of good nutrition and health behaviors. The trends were to be estimated from surveys prior to implementation of CBN (baseline), again one to two years after implementation (midline), and finally an end-line survey after five years. The target populations for inclusion in each survey were children under-3, for nutritional status, and mothers/caretakers of the children under-3, interviewed for an assessment of knowledge and practice of nutrition and health behaviors.

A two-stage cluster design was used for the surveys selecting clusters (census enumeration areas) with a Probability Proportional to Size from a listing, by tranche, of all the woredas combined. The survey teams visited Enumeration Areas (EAs) (or clusters) at baseline, and revisited these same EAs at midline and endline. Within each EA, between 15 and 18 households were systematically selected based upon presence of children under-3. The sample within each EA was redrawn at midline, rather than attempting to re-measure the same children. Sixty-five EAs in 50 tranche 2 woredas, and 65 EAs from 60 tranche 3 woredas, were surveyed at both base and mid-line. Both tranche 2 and tranche 3 baseline surveys collected data from all 65 EAs, but due to weather and logistical issues, only 60 EAs in tranche 2 and 61 EAs in tranche 3 were reached for repeated surveys at mid-line. For logistical reasons, the midline surveys were at the same time for both tranches, in September to October 2011.

In addition to household level surveys, a single EA level questionnaire was administered to either the HEW or a kebele representative on EA level characteristics (i.e. population, distance to nearest health center, etc) and CBN program components (i.e. number of HEWS, number of VCHWS, etc). Unfortunately, no EA level data was collected during data collection for the tranche 3 baseline, therefore no comparison can be made over time using EA level data.

B. Data analysis

Datasets were received from the separate firms tasked to collect data. Both baseline surveys were collected by Addis Continental Institute of Public Health and both midline surveys collected by Mela Research PLC.

i. Data cleaning

Individual and EA level data were entered into SPSS and analyzed using SPSS Version 16. Since baseline and midline surveys for each tranche were conducted by different organizations, initial data cleaning focused on reconciling variable definitions between baseline and midline datasets. Most importantly, values and labels defining the location and name of surveyed EAs (and subsequently woredas, zones, and regions) were reconciled between baseline and midline surveys in each tranche to ensure comparability. Additionally, all variables were investigated for data entry errors (values out of range, etc) and corrected where possible or set to missing.

During exploratory analysis of the baseline and midline individual datasets, extreme age heaping was discovered when investigating child age distribution. Bar graphs of age of child variables showed heaping particularly at 12 and 24 months. Age heaping is common occurrence in surveys since caregivers often report approximate estimates of the child's age in years as opposed to exact months. Child age is an important factor in the calculation of anthropometric outcomes as mis-reporting of age may result in incorrect z-score and under or over estimation of the prevalence of stunting and underweight. For this analysis, age heaping was particularly important as trends were to be estimated. This is because the differences that need to be detected over time by comparing two surveys may be as small as 1 ppts/yr, whereas comparing between, for example, regions from one cross-sectional survey, differences are much greater e.g. 5 ppts or more.

Extensive cleaning and re-calculation of the child age variable was conducted in all individual level datasets. In addition, variations on z-score variables (e.g. removing values for ages 12 and 24, and using non-exact ages) were explored to determine if heaping could be further reduced. No significant differences were found using variations on z-scores and the original calculations were used for analysis. This process is detailed in the Annex.

The newly created child age values were used to create updated z-score variables for height-for-age (HAZ) and weight-for-age (WAZ), and subsequently new prevalence estimates. The age adjustments resulted in a substantial change in prevalence estimates for stunting and underweight; actually increasing the point estimates at baseline and midline by between 2-7% in tranche 2 and 2-3% in tranche 3.

ii. Aggregating and combining datasets

Once cleaned, the baseline and midline individual level datasets were aggregated to the EA level, creating new datasets containing 60 (or 61 in tranche 3) cases representing each surveyed EA. The values within each case are thus no longer individual responses, but rather the mean or derived prevalence of all individual responses in each respective EA.

For each tranche, the aggregated baseline file was then merged with its respective aggregated midline file, creating a single file for each tranche containing 60 (or 61 in tranche 3) cases with two sets of values (estimates at baseline and midline) for each variable on household data. In addition, the midline EA level dataset for each tranche, derived from the EA level questionnaire, was then merged into its respective dataset so each case (EA) also included EA characteristic and program component data.

iii. Deriving program participation variables

Program participation variables were created to investigate associations between change in anthropometry and level of CBN participation (descriptions of derived variables can be found in table 2.1). A variable from the EA level questionnaire quantifying the length of time each EA had been conducting CBN activities was investigated. Unfortunately, it could not be used due to large inconsistencies in responses. CBN routine program data was thus used to calculate a "length of program activity" indicator, by subtracting the month in which routine data was first reported within each woreda containing the EA from October 2011 (the month of the midline evaluation surveys) in both tranches. It should be noted that routine data was reported at the woreda level but data was not available from kebeles (EA). Thus it is not possible to determine when the specific EAs in the evaluation surveys were included in the routine data reporting by the woreda. In other words, not all kebeles in each woreda began CBN activities and reporting at the same time. However, this estimate of "length of program activity" was the best available.

It was found that 17 EAs in tranche 3 were located in woredas that had yet to begin reporting any CBN routine data by the midline survey and were thus taken to be not yet implementing CBN (detailed later in Results section). A new variable was created coding EAs that had any months of reported data as one and those that had no reported months of data as zero. This was used to help explore associations with CBN implementation and change in outcome in tranche 3.

The second indicator to measure program participation was developed from a question in the EA level dataset on the length of time VCHWs spend per week on house-to-house visits, teaching, counseling, and community discussion. The values from each activity were summed to create a total VCHW activity time per week variable, which was then split into high (>7 hours) and low intensity (<7 hours) of activity based upon the distribution. This indicator proved telling in tranche 2, as all of the EAs surveyed had been implementing CBN for a minimum of 12 months (derived from the length of activity indicator). It was less telling for tranche 3, where some EAs had yet to begin CBN, and of those that had, a significant portion had been implementing CBN for less than 6 months. Therefore, the "intensity of CBN activity" indicator was used for tranche 2 only.

iv. Deriving outcome variables and estimating trends

All anthropometric values were converted to WHO Child Growth Standards using WHO Anthro Software and prevalence estimates created using -2 SDs as a cut-point. The original targets used in the PIM are in NCHS standards. All calculations here are in WHO Standards. For comparisons with targets in the discussion section, conversion to WHO Standards of the targets would make no difference to the specification of impact. For example, the target of stunting changing from 46% to 40% in 5 years, converted to WHO standards becomes 52% to 46%, both a six-percentage point difference[3] (Yang & de Onis, 2008).

[3] An excel program containing algorithms to convert NCHS estimates to WHO standards was used. This can be found at: [http://www.biomedcentral.com/content/supplementary/1471-2431-8-19-S2.xls]

To estimate trends in anthropometry, baseline and midline estimates of all anthropometric indicators were analyzed using paired-t tests at EA level to estimate the size and significance of any change between baseline and midline surveys.

IYCF indicators were created initially in the individual level datasets using the WHO indicator guide on IYCF practices (WHO, 2010a). Similar to estimating trends in anthropometry, paired t-tests were used to test for significant differences in baseline and midline prevalences of all IYCF indicators in both tranches at EA level.

Estimates of change in anthropometry and IYCF were also created by subtracting the aggregated midline estimates from the aggregated baseline estimates to create a difference variable for all z-score, anthropometric prevalence, and IYCF values. These results were, as expected, very similar to those estimated using individual level data. Definitions of outcome variables can be found in table 2.1.

v. Socioeconomic and environmental factors, HEP components, and other programs (TSF)

Changes in socioeconomic indicators, such as toilet facility, drinking water facilities, education, and roofing, were investigated using paired t-tests for significant changes over time. Additionally, these same indicators were categorized into improved versus non-improved/deteriorated, by creating difference variables (midline minus baseline) and then coding the difference values into one for improvement and zero for those that deteriorated/showed no improvement between baseline and midline. These ‘improved’ variables were used in regressions to control for potential confounding. Similarly, change in attendance of antenatal care (ANC) was investigated for change overtime as well by categorizing into improved versus unimproved.

The effects of participation in Targeted Supplementary Feeding program (TSF),a supplementary feeding program providing two three-months supplementary food rations targeted at moderately malnourished children under-5 and PLW, was investigated using several indicators. In the EA level questionnaire, HEWs were asked if their EA participated in “TSF, EOS only, or PSNP.” This question was incorrectly designed and did not allow for multiple answers, so participation in PSNP (Productive Safety Net Program – a food/cash-for-work initiative) was not able to be investigated. EA level participation in TSF, however, was available, although it is unclear if the responses were in reference to recent participation in TSF or any past participation, and was thus not used. Using a different approach a variable was created using data from the household questionnaires, concerning receipt of supplementary food for children attending the most recent TSF screening (where children under- 5 and PLW are screened for malnutrition using MUAC measurements). EAs were then categorized into those in which any children reported receiving food during the last screening and those with no children reporting receipt of food. This variable is then used as an indicator for participation in (or beneficiaries of) TSF. Definitions of all key variables can be found in table 2.1.

vi. Comparison data

The evaluation design was not able to include predetermined comparison groups, as discussed earlier, due to implementation priorities which required that the program be launched as fast as possible for all woredas in each tranche, with no deliberate phasing (that would have allowed for de facto comparisons). Therefore, trends in child growth outcomes are compared with historical trends from national surveys. Three DHS conducted in Ethiopia, dating back to 2000, allow for an 11-year estimate (2000-2011) of trends in anthropometry in terms of change in percentage points per year (ppts/yr) to compare with CBN evaluation trends. In tranche 3, delayed implementation (defined by no reporting of CBN routine data by the midline survey) does allow for de facto comparisons based on actual program implementation (reporting).

Additionally, CBN routine weighing data, collected monthly, is used to confirm trends. Trends from evaluation surveys and routine data are not directly comparable as program data is collected from *program participants* under the age of 2, while evaluation data was collected from randomly sampled *population* with children under the age of 3. However, the trends can be compared to confirm direction and rate of change.

C. Estimating changes in food security, seasonality, and exposure to drought

The impact of food security, seasonality, and drought on estimates of anthropometry and baseline and midline are evaluated to determine if the change between surveys can be linked change in environmental factors.

i. Food security

The household level questionnaire used for all surveys included a food security section. Within this is a three-question module constituting the Household Hunger Scale (HHS), an indicator used to measure household hunger in food insecure areas (Ballard et al, 2011), can be used to help assess the food security situation at both baseline and midline and monitor the change in reported food insecurity overtime.

Three categories of household food security are calculated from the HSS: little to no household hunger, moderate household hunger, and severe household hunger. All three categories are explored descriptively by comparing baseline and midline estimates in each tranche and checking for significant change overtime. Additionally, an indicator of change in little to no household hunger was created by subtracting the midline estimate from the baseline estimate in each tranche. This indicator is then used as a control variable to determine if improvement (or deterioration) in food security is related to improvement (or deterioration) in anthropometry.

ii. Seasonality

The timing of the baseline and midline surveys in each tranche was investigated to determine the respective season (hunger, harvest) each baseline and midline survey was conducted in. All surveyed woredas were mapped by region and compared with seasonality maps portraying location and timing of hunger and harvest seasons (USAID, 2010). Mapping was done separately for each region since seasonality differs in different regions of the country. The likely effects of seasonal differences between baseline and midline are estimated and used to interpret the extent to which seasonality affected the change in anthropometry seen in both tranches.

iii. Drought

Contextual data was used to determine the level of drought that surveyed areas experienced at both baseline and midline. Throughout each year, the Ethiopian Nutrition Coordination Unit (ENCU) assists in the classification of so called Hot Spot Priority woredas. Around 200 woredas are selected as Priority One (the most severe classification), based upon food security, food prices, and likelihood of drought. This classification was used to help determine the level of drought in surveyed areas at the time of the baseline and midline surveys in each tranche by cross-referencing the list of surveyed woredas with the list of Priority One woredas. A ratio was then created of the percent of woredas surveyed classified as Priority One in all available months surrounding the baseline and midline surveys. This ratio is then used to determine if drought (determined by the hot spot classification) was more prevalent in surveyed areas at baseline or midline.

3. RESULTS

A. CBN participation and implementation

The CBN was implemented in phases, with a different set of woredas – referred to as tranches – launched each year (as discussed above and shown in table 1.2). The baseline survey of tranche 1 was abandoned (for unclear reasons), so data from baseline and midline surveys on tranches 2 and 3 are used here. The baseline survey for tranche 2 was in June-July 2009, for tranche 3 in March-April 2010, and the mid-line survey for both in September-October 2011. Recall that baseline and midline surveys (and eventually endline) resample the same 60 EAs (clusters) by tranche, so the outcomes can be estimated as changes by EA, with n=~60 for each tranche; this use of 'repeated measures' increases the statistical power.

Implementation was expected to start in each woreda by tranche immediately after the baseline survey, which was timed to coincide with the CBN launch in that area. The first indication of progress in implementation came from the CBN routine data, as noted in the Methods section. These data came from monthly reports from the weighing component, made available by woreda, and showed actual launches by woreda from the timing given in these reports.

We first needed indicators of the extent of implementation—timing of start, number of woredas with program activities, participation within woredas, intensity of program from VCHW activities – both in order to describe program activities, and to link these (where they varied) to outcomes.

i. Length of time of CBN activity

The official launches of each tranche were around July-August each year, with HEW and VCHW training by woreda. Training of HEWs was completed prior to launch. However, the start of the CBN activities did not always immediately follow the launch in all woredas (as we will see below).

Questions on the length of time each EA had been conducting CBN activities were included in the EA-level questionnaire, addressed to the HEW. Unfortunately, due to many data entry errors in the necessary variables, that information was unavailable for analysis, as discussed in Methods section.

Routine CBN weighing program data by woreda (not EA) was therefore used as an alternative. From this, a "length of program activity" indicator was calculated by subtracting the month in which routine data was first reported within each woreda from October 2011 (the month of the midline evaluation surveys), for both tranches. Routine data reporting was at the woreda level, and thus could not be analyzed by EA. Not all EAs in each woreda began CBN activities and reporting at the same time. However, is considered to provide a reasonable estimate of length of program activity.

Estimates on the length of program activity in each tranche are shown in figure 3.1. Tranche 2 was launched in July 2008 (28 months before the mid-line survey). The largest number started reporting in January 2010, at 21 months before the mid-line survey. In tranche 3, 17 woredas, or 28%, had yet to begin reporting CBN routine data at all by the midline survey, and thus were taken to not have begun implementing CBN. Of those EAs in woredas that had begun reporting data, 32% had been reporting for over six months. This measure not only gives an indication of the rate of launching, but allows an estimate of program implementation, most importantly for tranche 3, where those woredas that had not reported starting may need to be treated differently, as they can be expected not to show impact of the CBN; they thus also provide for a *de facto* comparison group. Further descriptive data on EAs and length of program activity in each tranche can be found in table 3.1.

The "length of program activity" in tranche 2 can be used to investigate whether there are associations of HEW and VCHW activities, with changes in anthropometry. In tranche 3, EAs in woredas with no routine data can be compared to the others to investigate if the presence of CBN is associated with change in anthropometry. These results are presented later in section 3 B.

ii. *Training in CBN*

Training received by frontline workers was assessed at the mid-line survey, see table 3.2 (this was also assessed at the baseline evaluation survey for tranche 2, data not shown). Around 90% of EAs in both tranches reported that 100% of HEWs working had been trained in CBN, with just over 80% of EAs in both tranches also reporting HEWs had received refresher training. HEWs were also reported to have received additional training in Essential Nutrition Actions (ENA) in 55% of EAs in tranche 2 and 43% of EAs in tranche 3. Several EAs in both tranches also reported HEWs receiving additional training in Enhanced Outreach Strategy (EOS), additional food support, and safety net programs. It was not possible to determine clearly, the percent of VCHWs *initially* trained in CBN due to data entry errors.

By midline, <u>82% of EAs in tranche 2 and 77% of EAs in tranche 3 reported that VCHWs had received *refresher* training in CBN.</u>

iii. *Participation in CBN*

Participation in child weighing, as reported by caregivers, within three months prior to survey, was also estimated to indicate participation in CBN. The percentage of children in the samples reported weighed (shown in table 3.3, tranches 2 and 3, first line) increased from 14% to 33% ($p<0.001$) between baseline and midline in tranche 2. A similar increase is seen in tranche 3, with the percentage of children reported weighed more than tripling from 7% to 30% ($p<0.001$). This includes both health facility and community weighing. Weighing at the health facility usually occurs following referral by the HEW to the local health facility. Community weighing is usually conducted by the VCHW as part of routine program activities. Any reported weighing, no matter the modality, is used as a measure of participation in CBN, as both health facility and community weighing involve contact with either the HEW or VCHW.

Of those children reported as weighed, the percent of those weighed who were weighed at a health facility decreased in both tranches from around 80% to around 50% ($p<0.001$). The percent of children weighed who were weighed at a community weighing session increased from 22% to 51% in tranche 2, and from 11% to 55% in tranche 3 ($p<0.001$ for changes in both tranches). Thus there was a substantial increase in the number of children weighed, and a major shift towards weighing at community weighing sessions.

Total weighing at health facilities increased slightly (by around 4 percentage points (ppts), not significant: table 3.3, fifth line) in both tranches between baseline and midline. The proportion of the total population participating (in the last 3 months) in community weighing sessions increased from 3% to 16% in tranche 2 and from 1% to 19% in tranche 3 ($p<0.001$).

These estimates give an idea of <u>the likely extent of participation of children (in the previous three months) by CBN: just less than 20%, from the community weighing data; around 30% from data on total weighings, including at a health facility.</u> This is averaged over all EAs for tranche 3, while we think that 28% of EAs may not have started. Selecting the 72% of EAs that had started in tranche 3, the midline levels of weighing were similar to tranche 2, 31%, and of percent population at community weighing session were 17%.

iv. Participation by child's age

Although CBN is designed to target children under-2, a large number of children aged 24 to 35 months reported participating in weighing. The data can be broken down by age band from the known child's age. Table 3.4 gives the percent of children within each age group (0-11 months, 12-23 months, and 24-35 months) reporting participation in weighing in the three months prior to the baseline and midline surveys. Children between 12 and 23 months reported participation more often than children 0-11 months or 24-35 months at midline. In tranche 2, children 0-11 months reported the lowest participation at midline at 23%, compared to 41% of 12-23 month olds and 37% of 24-35 month olds. In tranche 3, 30% of children aged 0-11 months participated at midline, compared to 36% of 12-23 month olds and 27% of 24-35 month olds. At midline in both tranches, children 0-11 months were least often weighed at community weighing session.

Through time, all age groups experienced significant ($p<0.05$) increases in any participation between baseline and midline. More specifically, reported participation in community weighing increased significantly among all age groups ($p<0.001$), while changes reported participation in health facility weighing were less substantial and not generally significant.

This result should <u>prompt a reconsideration of program guidelines</u>: is it necessary (or feasible, or desirable?) to reinforce the aim of restricting CBN participation to children under-2? (Note from an evaluation viewpoint it is useful to keep 2-3 year-olds in the sample, as the effects are intended to persist.)

v. Possession of Family Health Card

Family Health Cards (FHC) are used in CBN to track the growth of the child at monthly weighings and act as visual aids for the HEW to communicate about child growth trends to caregivers. Possession of an FHC can be used as an indicator of contact with either an HEW or VCHW and hence participation in CBN. Since NGOs and other organizations also conduct growth monitoring programs, respondents were asked if they had a FHC (used for CBN), any Other Health Card (OHC) (used for a different program), or both. Results are given in table 3.5. The percent of population with any type of health card (FHC, OHC, or both) increased (though not significantly) from 50% to 64% in tranche 2, and from 46% to 61% ($p<0.001$) in tranche 3.

Of those reported to have any type of health card, the percent with only an FHC increased from 22% to 41% ($p<0.001$) in tranche 2, and (not significantly) from 30% to 35% in tranche 3. Subsequently, there was a decrease in the percent reporting possession of only an OHC in both tranches, but particularly in tranche 2, with a reported decrease from 68% to 31% ($p<0.001$). From the total population, at baseline in both tranches, the percent of people reporting possessing only an OHC was much larger than those reporting possessing only an FHC (33% versus 11% in tranche 2, and 23% versus 14% in tranche 3). At midline, though, the percent of the total population reporting possession of only an FHC surpassed those with an OHC only as well as those with both in both tranches.

The estimate of 61-64% of respondents having health cards gives a further estimate of program participation. This is higher than the 20-30% participation in the last 3 months estimated from weighing, as expected. The significant increases, in both tranches, by about 15 ppts from base- to mid-line (table 3.5, first line) suggests an impact of CBN on health card use; and other data in table 3.5 suggest this involved a shift towards Family Health Cards.

vi. Contact with HEWs and VCHWs

An increase in the reported contact with both HEWs and VCHWs was found in both tranches, as seen in tables 3.6 and 3.7, respectively. The percent of population with no reported contact with an HEW in the previous six months decreased significantly in both tranches between baseline and midline, as the percent with 4-6 contacts with an HEW increased significantly from 1% to 28% in tranche 2 and 18 % to 34% in tranche 3 ($p<0.001$ for changes in both tranches). Additionally, reported contact with an HEW greater than 6 times in the previous six months increased to above 10% in both tranches (from 0% in tranche 2 and 3.0% in tranche 3).

HEWs: The highest reported location of contact with an HEW at midline was a health post in both tranches, but the number of people meeting HEWs at community outreach, house visits, community conversation, growth-monitoring, and model family training all increased significantly ($p<0.05$). At least one contact with an HEW in the previous 6 months was reported at 85% at midline, and more than three contacts at about 40% in both tranches. There was also a significant increase ($p<0.001$) in the reported receipt of information on child weight/growth, complementary feeding, family planning, and child caring practices from HEWs in both tranches.

VCHWs: The percent of the population (see table 3.7) reporting no contact with VCHWs in the 6 months prior to the survey decreased significantly ($p<0.001$) by around 20 ppts in tranche 2 and 30 ppts in tranche 3 (i.e. to about one third with no contact), while the percent that reported contact 4-6 or more times increased from 20% to 35% in tranche 2 and from 10% to 30% in tranche 3 ($p<0.001$ for changes in both tranches).

There was a significant increase in reported contact with VCHWs at community outreach, house visits, community conversation, as well as growth monitoring in both tranches; with the highest reported location being house visits at midline in both tranches. With the exception of family planning in tranche 2, significant increases ($p<0.001$) in reported information received on child weight/growth, exclusive breastfeeding, complementary feeding, and child caring practices from VCHWs was found between baseline and midline in both tranches.

The data on both HEW and VCHW activities suggest sizeable and significant increases in contacts and outreach between baseline and midline. In CBN areas it seems that about two-thirds of the population had some contact with these frontline workers by mid-line, and at least one third had regular contact (more than 4 times in 6 months), and with increased frequency.

vii. *Summary of indicators of program coverage*

Key questions concern how extensively CBN was implemented, and what percentage of children benefitted from CBN activities. A summary of the available indicators is given in table 3.8, extracted from the results discussed earlier in this section.

Training of VCHWs – refresher training was estimated – appears to have been extensively carried out, covering about 80% of VCHWs in both tranches by the midline survey. The next five indicators, all of which increased significantly by midline, give different ways of assessing program participation. Child participation in weighing is included as this in practice is a basis and entry to counseling, and possibly referral[4]. Weighing can be at the facility, often as a result of referral. So community weighing may give the lowest estimate of recent participation. Contact with the VCHW in the last 6 months is another estimate of participation (and is in line with the expectation that not every contact is at weighing), and the estimate of about one third of children contacting more than three times may be a reasonable basis for estimating that about one third of children are active participants in the CBN. (It also happens to be almost exactly in line with the estimate we made on quite different grounds, from the routine weighing program data (Hoblitt & Mason, 2010; Buback & Mason, 2011; White & Mason 2011).

viii. *Estimates of program intensity*

As noted earlier (figure 3.1 and table 3.1) 28% (17/61) of the tranche 3 EAs had not yet begun implementing CBN by the time of the midline survey. Therefore in many analyses, the 17 – not started – EAs were excluded, and could also be used as a comparison group. In contrast, the program in tranche 2 had been implementing for considerably longer and all EAs had begun activities by the midline survey. These results are included in table 3.9.

A number of indicators were explored to estimate the level of program activity (or intensity). In tranche 2 (see Methods section) the indicator arrived at was the number of hours spend by VCHWs spent weekly on CBN activities (see figure 3.2); this was dichotomized at 7 hours (approximately the mean) and the EAs with greater than 7 hours spent by VCHWs were defined as having a higher intensity of program activity. Of the EAs in tranche 2, 52% fell into this category. In contrast, for tranche 3, the program had been operating for a much shorter time,

[4] Hopefully we do not need to go into the self-evident point that the act of weighing itself does not cause a child to grow. Weighing may be seen as analogous to taking a child's temperature – it does not affect a fever, but is a really important piece of information in deciding what to do.

and in those 44 EAs which reported program activity, 50% had reached the level of 7 hours or greater per week. No association of this variable with outcome was found in tranche 3, whereas it was significantly usefully associated in tranche 2, as will be discussed later.

An indicator used elsewhere (WHO, 2012) is the ratio of VCHWs to children, with 10-20 children per VCHW being considered well in the range of that likely to produce impact. The indicator of percentage of EAs with an average of less than 20 children under-2 per VCHW is shown in table 3.9, reaching 52-58% in the two tranches at midline. In tranche 2, the average VCHW intensity for all EAs was 1:18 (or 1 VCHW per 18 children under-2). In tranche 3, for all *reporting* EAs, the average VCHW intensity was 1:28, although this average is inclusive of two extreme outliers (two cases had VCHW intensities around 1:177). If the two outliers are removed, the VCHW intensity in *reporting* tranche 3 EAs decreases to 1:19. Both estimates of VCHW intensity are in line with the range expected to produce the most impact.

B. Changes in outcomes between base- and mid-line surveys

i. Infant and young child feeding (IYCF)

Indicators of infant and young child feeding were derived from the individual caretaker questionnaires at baseline and midline. A summary of the most important results is in table 3.10. The most striking effects were on exclusive breastfeeding under six months, minimum dietary diversity, and minimum acceptable diet between 6 and 23 months of age.

Exclusive breastfeeding rates in tranche 2 significantly increased ($p<0.001$) by 20 percentage points (ppts) to 89%, while tranche 3 decreased by 10 ppts to 78%, although the change is not significant. No change was seen in either early initiation of breastfeeding or continued breastfeeding at 1 year – which is anyway above 90% – between baseline and midline in either tranche.

Minimum dietary diversity in children 6-23 months, defined as the child receiving food from four or more food groups in the previous day (table 3.10, fifth line), showed a large and highly significant improvement between baseline and midline, of around 20 ppts in tranche 2 and around 15 ppts in tranche 3, raising the percent of children meeting minimum dietary diversity at midline to around 50% in both tranches ($p<0.001$). This significant improvement is also seen within narrower age bands of 6-11 months, 12-17 months, and 18-23.

Figures 3.3 and 3.4 show the distribution of children who received food from each of the seven different food groups during the previous day. Dairy product consumption increases in both tranches ($p<0.10$), as does flesh foods [meat] ($p<0.05$), and vitamin A rich fruits and vegetables ($p<0.05$). Tranche 2 also showed a significant ($p<0.05$) increase in the consumption of eggs.

Minimum meal frequency, defined as consuming solid, semi-solid, or soft foods the minimum numbers of times appropriate for each age group[5] improved significantly in breastfed children in

[5] Breastfed children: 6-8 months = 2 times per day, 9-23 months = 3 times per day. Non-Breastfed children: 6-23 months = 4 times per day (WHO, 2010b).

tranche 3 ($p<0.05$) from 58% to 66%. While no change was seen in tranche 2, the percent of children meeting minimum meal frequency was still high, also around 67%. Non-breastfed children were not explored here because the sample size of non-breastfed children consisted of only 12 cases.

The percent of breastfed children consuming a minimum acceptable diet, defined as meeting both the minimum dietary diversity and minimum meal frequency, almost doubled, increasing prevalence of minimum acceptable diet to 44% in tranche 2 ($p<0.001$), and increased by 10ppts to 37% in tranche 3 ($p<0.01$). The prevalence was not able to be determined for non-breastfed children at baseline or midline due a lack of data.

<u>These changes show a considerable improvement in child feeding practices</u>, along the lines that HEWs and VCHWs are trained to provide counseling on. The intent in these sections of the training manuals is being met, but there is still some way to go: 37-43% of 6-23 month children are now getting a minimum acceptable diet (meal frequency and 4 food groups). This has been significantly improved by the CBN. <u>But more than half the children in the program EAs are still not getting a minimum acceptable diet</u>.

ii. Care of childhood illness

The frontline workers are trained to counsel on improving hygienic habits and facilities, for which reduction in diarrhea incidence is a primary aim. They also aim to improve the care of children who do get diarrhea. Frequency of diarrhea and its treatment were assessed, as shown in table 3.11.

Reported *prevalence of diarrhea* in the two weeks preceding the survey decreased significantly by 10 ppts to 21% in tranche 2 ($p<0.001$) and 5 ppts to 22% in tranche 3 ($p<0.05$). This matches improvements in sanitary facilities, as discussed later in section 3 C and table 3.18. Behaviors aimed at *preventing* diarrhea were not assessed.

Treatment of diarrhea (assessed among those reporting diarrhea in the previous two weeks) improved in both tranches, in certain aspects. The percent of caregivers that reported giving their child oral rehydration salts (ORS) during their last episode of diarrhea increased significantly in both tranches by 10 ppts or more ($p<0.05$), and the practice of reducing fluid intake during diarrhea episodes was significantly reduced, although still unfortunately prevalent (in around 30% of cases at midline in both tranches). Concomitantly, giving the same fluids as usual increased to 50% in both tranches, but the message to *increase* fluids during diarrhea – even though ORS are quite common – has not been successfully put across.

Feeding practices during the diarrhea episode improved somewhat, with some increases reported in food given. An increase in caregivers giving the child about the same amount of food as usual was found. Similarly, there was a significant decrease from 55% to 30% in tranche 2 and from 55% to 28% in tranche 3 in children reported to have received much less than the usual amount of food during their last illness ($p<0.001$ in both tranches). However, only 20-30 % of caregivers continue with usual amounts of food at midline. Here too <u>messages on treatment of diarrhea are not appearing altogether effective.</u>

iii. *Assessing nutritional status of children and interpreting trends*

Improvement in child nutritional status is the primary impact objective of the CBN program, and of the overall National Nutrition Strategy (NNS) (FMOH, 2008, p 12 and 93). Specifically, the first two impact indicators are, by 2013: [6]

- To reduce underweight from 38% to 30% (2 ppts/yr)
- And stunting from 46% to 40%, (1.5 ppts/yr)

The program started in 2008, implying the rates shown for underweight and stunting for five years up to 2013. (As seen below, these are not far different from the expected without-program secular trends.) Wasting, the third objective, is unstable and affected by proximal factors (usually current illness) and intermittent drought; wasting is used here as a secondary (and less understood) indicator. The fourth impact indicator is low birth weight, which there is no currently feasible means of assessing with any accuracy. The evaluation therefore gave high priority to assessing changes in anthropometric indicators, in line with the project's primary impact objectives, and stunting proved the most valuable indicator.

Assessment of whether the with-program rates are significantly higher than without-program was complicated by the impossibility of surveying non-CBN areas: government policy was that programs should start in all areas (within tranches) immediately, and no non-program area could be surveyed. The original intent was to use the trend between base-line surveys, allowing for differences by area, as a measure of the non-program trend. This was undermined by the abandonment of the baseline survey for tranche 1. Therefore, three approaches were relied upon for interpreting trends:

1. Comparing with-program trends with those estimated from DHS and with CBN routine data;
2. Using variations in program delivery (available in tranche 2) to investigate associations with anthropometric changes;
3. In tranche 3, some 30% of woredas had not reported any program activity and were taken as non-implementing: comparisons between those implementing and not facilitated interpretation.

Trends in stunting, underweight, and wasting from three DHS surveys between 2000 and 2011 are shown in table 3.12. Changes in percentage points per year (ppts/yr) are shown to be quite stable, at around 1.3 ppts/yr improvement in stunting, and 0.7-1.9 ppts/yr in underweight. We should note that the secular stunting rate is almost the same as in the program objective referred to above. In a broader context, these rates are rapid for African countries, where 0.2 ppts/yr is the average for underweight in East Africa, and 0.1 ppts/yr in stunting (UNSCN, 2010, p 47-48). So the underlying (no-program) trend in Ethiopia is already quite rapid. Any change greater than 2 ppts/yr would be unlikely without effective intervention (the record for underweight improvement is 2.7 ppts/yr, briefly seen in Thailand in the 1980's, from ACC/SCN, 1996).

The CBN routine data from the weighing program became available between the base- and mid-lines, and an example of the pattern of underweight through time has been introduced in figure 1.1. These data suggested that the trend in the first 1-2 years after launch was about 10 ppts/yr improvement in underweight prevalence, *among participants*. The participation rate was

[6] In NCHS standards

estimated at about 30%. Thus *at population* level (as in the surveys) we would expect about one third of the rate, about 3.3 ppts/year (Hoblitt & Mason, 2010; Buback & Mason, 2011; White & Mason 2011). This usefully gives an order of magnitude expectation of what the evaluation surveys might show in underweight trends. Stunting rates would be expected to be similar.

From the evaluation surveys, anthropometric measurements (involving weight, height, and age) are calculated for individual children as a standard deviation or z-score, compared to WHO Child Growth Standards. The mean z-score, aggregating by EA, is the basic measure, assessed on the same EAs (but with a different sample of children) at both baseline and midline. The difference for each EA is then also calculated. The z-score is usually the first index looked at, containing more information than when dichotomized (at – 2 SDs) to give prevalences of stunting, underweight, and wasting. However, prevalences are more readily available and more easily understood. So both indicators are used here.

Estimates of changes between base- and mid-line surveys are summarized in table 3.13a and 3.13b, and discussed below.

iv. Stunting

A significant improvement in height-for-age z-scores (HAZ; of 0.26 z-score units: -1.731 to -1.473; p<0.05)[7], was seen in tranche 2 between baseline and midline surveys (see table 3.13a). Improvement in HAZ was also seen found in Tranche 3 overall (with p=0.1) (table 13b). These differences in stunting prevalence in tranche 2 were from 50.5% to 40.6% (9.9 ppts; p<0.001), and from 42.9% to 38.5% (4.4 ppts; p=0.11) in tranche 3.

However, as discussed earlier, 17 of the 61 EAs in tranche 3 had not begun reporting routinely (i.e. as seen from the CBN routine datasets), so presumably had little or no CBN implementation. Comparing these (tables 3.14a and 3.14b) with reporting EAs shows that the non-reporting EAs had some increase in stunting, whereas the reporting EAs had a decrease in prevalence from 46.3% to 38.3% (8.0ppts, p<0.05), with associated improvements in HAZ. The improvements in stunting further analyzed here for tranche 3 are those for reporting EAs. The non-reporting EAs provide for a *de facto* without-program comparison group.

Confidence intervals (95%) are shown in tables 3.13a/b and 3.14a/b. These are in line with the significance estimates: where the range does not include zero the difference is significantly (p<0.05) different from zero. The expected without-program rate is -1.3 ppts/yr, so for tranche 2 over 2.3 years this is -3.0 ppts. This is outside the CI range (-14.8 to -4.9 ppts, mean -9.9), thus the difference in tranche 2 is significantly greater than the expected without-program difference. For tranche 3 (reporting EAs only) the difference is -8.0 ppts, 95% CIs -14.7 to -1.2 (table 3.14a). This includes the -2.2 ppts (-1.3*1.5) expected without-program; the 90% CIs lower bound is -2.5 (excluding -2.2), thus the significance is between 0.1 and 0.05 that the tranche 3 difference is higher than the long term trend. Given that the actual period of implementation has a median of 6 months, the calculation based on 1.5 years is very conservative. In any event, the trend in tranche 3 is likely to be greater than the expected without program rate with a p<0.1.

[7] As a rule of thumb, 0.1 z-score units change in HAZ or WAZ usually equals 4ppts of prevalence

For standardizing the comparison with expected without-program trends, the stunting differences can usefully be translated into ppts/yr. The denominator used is the period between baseline and midline surveys, 2.3 years (28 months) for tranche 2 and 1.5 years (18 months) for tranche 3. The median periods of actual implementation (see Figure 3.1) were 21 and 6 months respectively. Therefore this again gives a conservative estimate of the rate, especially in tranche3. The rates using the longer denominator represent those achieved from the official program launch.

The estimated rates of stunting decrease are -4.3 ppts/yr in tranche 2, and -5.3 ppts/in tranche 3 (reporting EAs). These improving rates of stunting are shown in Figure 3.5 – 3.7, compared to the long term trend (dashed line). <u>The rate of improvement within CBN EAs is thus much greater (by 3-5 ppts/yr) than the expected improvement of -1.3ppts/yr based on the 11-year DHS historical trend</u> (between 2000 and 2011). It is also similar to the rate of around 3.3 ppts/yr expected for the whole population from the CBN routine data results on participants only, with 30-40% participation rate.

v. *Underweight and wasting*

Changes in weight-for-age and underweight between baseline and midline were not significant, either as z-scores or prevalences (see table 3.13a/b). Given the high association between height- and weight-for-age, this is presumably related to worsening of weight-for-height z-scores (although this does not always show up in wasting prevalences).

Wasting changes significantly for tranche 2, increasing from 9 to 15% between surveys. Drought was more severe in 2011 than 2009 (FAO, Crop prospects, Dec 2011, p 14). It is possible that the tranche 2 woredas were more affected than tranche 3 (but see later on hotspots in section 3 E). In any event, it is likely that the reason that underweight prevalences do not change in tranche 2 in line with stunting is that wasting increased. In tranche 3 the story on underweight and wasting, related to stunting, is not so clear.

vi. *With-program stunting trends compared to varying program implementation*

Although there could be no formal comparison groups, advantage can be taken of opportunities from *de facto* (and unplanned) differences in implementation giving internal comparison groups, to investigate whether greater implementation is associated with more improvement in anthropometry. Two opportunities arise:

1. In tranche 2, varying resource commitment to the program (intensity) can be indicated by hours spent by VCHWs conducting CBN activities each week; this variable is dichotomized into EAs with high and low intensity (in terms of hours spent on CBN activities) and changes in stunting are compared between the two groups;
2. In tranche 3, 17 woredas did not begin reporting on the weighing program, as discussed above, and are considered likely to have had limited program implementation: thus reporting and non-reporting EAs can be compared.

COMPARISON OF STUNTING TRENDS WITH INTENSITY IN TRANCHE 2.

Those EAs with higher intensity of CBN activities had significantly greater reduction in stunting prevalence – 54.5% to 38.8 % (15.7 ppts) – compared to the low activity group – 44.0% to 41.2 (2.8 ppts); the difference in stunting between high and low activity was 12.8 ppts[8]. Results are shown in table 3.14c/d). HAZ results were similar but less significant (p=0.12).

These results are shown in table 3.15 (first 4 rows, tranche 2). The effect size on stunting can be estimated as the difference between high and low activity, i.e. 15.6-2.8=12.8 ppts, as given above. In the regression analyses discussed in detail later, the coefficients controlling for a range of potential confounders are similar, shown in the last row of table 3.15. The net effect size is estimated at 11.2 ppts.

This association helps establish that level of CBN program activity is significantly related to improved child nutrition.

COMPARISON OF STUNTING TRENDS WITH PROGRAM PARTICIPATION IN TRANCHE 3.

Overall, tranche 3 was implemented for a shorter time than tranche 2 – see figure 3.1 – and nearly 30% of woredas did not report activities at all. The comparison is therefore made between EAs in these non-reporting woredas and EAs in reporting/implementing woredas. Results are shown in table 3.14a/b. The difference in change of stunting between implementing and non-implementing woredas was 12.4ppts in stunting – very similar to the 12.8 ppts in tranche 2 – and 0.58 z-score; both these are significant (p=0.046 and 0.003 respectively).

Controlling for potential confounders, see table 3.15, actually increases the estimated effect size (of implementing vs. not) to 13.8 ppts.

This result adds to the evidence that program participation is associated significantly with reduced stunting.

C. Changes in socioeconomic and environmental factors

The bivariate associations established above cannot however lead to deciding on possible causal effects of the CBN until alternative explanations are addressed – i.e. potential confounders taken into account. Specifically, the improvements in child nutritional status might be associated with *other* factors, such as education level, or *change* in education level. Do EAs with improving stunting have better education; or did education improve more in these than others during implementation, which could account for the observed significant stunting changes? To address this we need to examine socio-economic and environmental factors, and participation in other programs. In this section, the levels and changes in these variables are discussed. Controlling for these in relation to anthropometric outcomes is discussed in a later section.

[8] This difference, between group mean stunting differences (i.e. 15.7 or 2.8) was also significant, p= 0.02

i. Education

The prevalence of overall poor education (defined as having no schooling) in caregivers, shown in table 3.16, decreased from 75% to 71% in tranche 2 (although not significant) and from 74% to 65% in tranche 3 ($p<0.001$). Within this there were only small changes in primary education, secondary education, etc. This may not be actual additional education itself, as caregivers are mostly beyond school age, certainly primary school. Thus these results may show some differences due to resampling within EAs between base- and mid-line. In tranche 3 where the difference is significant, this may indicate a somewhat better off sample redrawn at midline, indicating that we need to control for education.

The 2011 Ethiopian DHS (EDHS) reported a 52.1% total prevalence (58.1% rural) of no education in females (CSA, ICF International, 2012). Therefore, while education has improved, prevalence of poor education at 65-70% at midline is still higher in surveyed CBN areas than the national average.

ii. Roofing

The percent of the population with poor roofing is used as a proxy for poor socioeconomic status (SES). In both tranches – see table 3.17 – the percent of the population with poor roofing material (anything other than corrugated iron sheeting), decreased significantly by 12 ppts in both tranches ($p<0.001$). The largest changes appeared to be a significant ($p<0.01$) decrease in the percent of the population with thatch or grass roofs in both tranches, and a subsequent significant ($p<0.001$) increase in the percent of the population with corrugated iron sheet from 31% to 43% in tranche 2 and from 33% to 45% in tranche 3. Overall, the prevalence of poor roofing in tranche 2 decreased from 69% to 57% ($p<0.001$) and from 67% to 55% in tranche 3 ($p<0.001$).

This may well represent real improvement in SES between baseline and midlines.

iii. Toilet facilities

The percent of the population using poor toilet facilities (as defined by DHS) significantly decreased ($p<0.001$) from 93% to 77% in tranche 2 and from 96% to 82% in tranche 3, see table 3.18. The largest change reported between baseline and midline was an increase in reported use of pit latrine with slab, from 1% to 19% in tranche 2 ($p<0.001$) and from 1% to 10% in tranche 3 ($p<0.001$). There was also a large reported decrease in use of no facilities/bush in tranche 2 from 28% to 16% ($p<0.001$), and a significant ($p<0.01$) decrease in disposing on farm in both tranches. The CBN surveyed EAs appear to have started with worse than improved reported toilet facility usage but improved to be better than the national average of 84.5% (90% rural, from EDHS 2011) by midline.

<u>Upgrading toilet facilities is one of the important aims of the HEP and CBN, thus the increase in improved facilities - e.g. from 7 to 23% - is good news for these programs.</u> For the analysis here, controlling for toilet facilities is needed, as its relation to nutritional status has been shown in many analyses.

iv. Source of drinking water

The percent of the population using poor drinking water sources (as defined by DHS) decreased in tranche 2 from 40% to 34% (although not significant), and experienced little change in tranche 3, remaining around 45% at both baseline and midline (see table 3.18). While there were some decreases in utilization of certain 'improved' water sources (protected well and tube borehole), the largest change in both tranches appears to have been a significant increase in the percent of people using a public tab or standpipe as a drinking water source from 18% to 35% ($p=0.001$) in tranche 2 and 18% to 26% ($p<0.05$) in tranche 3. Significant decreases were also seen in the percentage of people using unprotected wells, down 2 ppts, and unprotected springs, down 7 ppts, as their drinking water sources in tranche 2. Compared to the 2011 EDHS reported 48.9% (58.1% rural) total poor drinking water sources, tranche 2 surveyed EAs have better than the national average, while tranche 3 surveys EAs have remained similar.

As for toilet facilities, these results are good news for the overall health improvement programs. And again, for our purposes, we need to control for water supply, and have the necessary variables.

D. Associated programs: maternal health and Targeted Supplementary Feeding

i. Antenatal Care

Reported attendance of any ANC increased significantly in tranche 2 from 54% to 69% ($p<0.001$) and from 56% to 62% in tranche 3 ($p<0.05$) (table 3.20). While any reported attendance at ANC increased, there was little change in the number women reporting attending ANC either two to three or four or more times between baseline and midline surveys in either tranche. However, the percent of women that reported attending ANC four or more times in CBN surveyed areas at midline (35-40%) is still higher than the national 2011 EDHS estimate of 19.1% (14.4% rural). ANC was also most often reported to have taken place at either a health center or health post.

In addition to the reported increase in ANC attendance, a significant increase in receipt recommended ANC services was found, as seen in table 3.21. Significant increases were seen in reported weighing, urine testing, and blood testing at ANC. Additionally, there was a significant increase ($p<0.01$) in the percent of women reporting receipt of counseling on breastfeeding and HIV/AIDS during ANC in both tranches. Receipt of counseling on family planning and maternal nutrition significantly increased ($p<0.01$) in tranche 2.

These results suggest considerable effectiveness of the HEP and CBN in improving access to and use of antenatal care; of the services provided (e.g. tests and malaria prophylaxis); and of the counseling.

ii. Health care and nutrition during pregnancy

Data on intake of iron supplements, tetanus immunization, and de-worming during pregnancy are shown in table 3.22. Reported intake of iron/folate during the most recent pregnancy increased significantly from 29% to 50% in tranche 2 ($p<0.001$), and from 31% to 39% in tranche 3 ($p<0.05$). The reported length of time iron/folate was consumed also increased between baseline and midline, particularly in tranche 2 at baseline, where 100% of the women consuming iron/folate reported taking it for only between 1 and 7 days. This increased impressively by the midline survey in tranche 2, with 39% of women reporting 16-30 days, and 18% of women reported taking iron/folate for 31 or more days. A similar shift was seen in tranche 3 data.

A significant ($p<0.05$) decrease in reported receipt of any tetanus injections during last pregnancy was estimated in both tranches, although there was a shift to more injections received by one person, with 70% reporting having received 2-3 injections. Tranche 3 also showed a decrease in number of women immunized. The reasons for this decrease, which is against the trend of other interventions, needs further research. It might be because most women had already been immunized in the previous (first?) pregnancy and the policy was not to repeat this; it could be due to supply shortages; or other reasons.

The reported receipt of a de-worming tablet during the previous pregnancy remained unchanged between baseline and midline in tranche 2, and experienced a slight decrease from 12% to 11% ($p<0.01$) in tranche 3. This is again out of line with the improvement in other interventions, and could indicate supply constraints.

At baseline in both tranche 2 and 3, over 60% of caregivers had reported consuming *less* food than usual during their last pregnancy, as seen in table 3.23. This presumably reflects the custom of 'eating down' during pregnancy with the aim of avoiding a difficult delivery; changing this habit is an objective of the counseling provided. At midline, a significant decrease in reported reduced consumption during pregnancy was found in both tranches, down to 47% from 64% in tranche 2 ($p<0.005$) and to 55% from 63% in tranche 3 ($p<0.01$). In accordance with this, more mothers continued with the same intake as usual, increasing from 27% to 45% ($p<0.005$) in tranche 2, and from 31% to 40% ($p<0.01$) in tranche 3.

<u>This positive trend in food consumption and use of iron-folate during pregnancy indicates effective counseling before and during pregnancy, by HEWs and VCHWs. Such successes could be emphasized during initial and refresher training.</u>

iii. Delivery care

A small increase in reported institutional delivery was found in tranche 2, from 4% to 7% ($p<0.05$), but the incidence remains very low (see table 3.24). Home delivery is the norm, and the assistance at home delivery is thus a key factor. Over 90% of assistance is provided by a traditional birth attendant or relative/friend, and this is hardly changing. Within this, there is some shift away from TBAs, and some increase in assistance from health workers. Overall however this aspect of maternal care is changing little. Weighing at birth, while increasing, remains below 20%.

iv. Post-delivery care

The percent of women that reported receiving a post-natal (or post-delivery) visit increased significantly in tranche 2 from 22% to 30% ($p<0.01$), and from 18% to 29% (although un-significant) in tranche 3 (table 3.25). Timing of the post-natal visits also improved, particularly in tranche 2, with visits within 1 hour of birth increasing from 9% to 28% ($p<0.001$) and visits more than 48 hours after birth decreasing from 72% to 34% ($p<0.001$) between baseline and midline. Similar changes are seen in tranche 3, although not significant. Further indication that post-natal care is improving is the significant increase in the number of women reporting to have received a Vitamin A supplement within two months of birth in both tranches, increasing from 24% to 40% in tranche 2 and 19% to 30% in tranche 3 ($p<0.001$ for both changes). HEWs were reported as those most often conducting post-natal visits at both baseline and midline in both tranches.

v. Targeted Supplementary Feeding

Targeted Supplementary Feeding (TSF) is a food distribution program, targeted at food insecure areas, that provides moderately malnourished children and PLW with three-month rations of food, following bi-annual or quarterly MUAC screenings. The entry into the program is screening (by MUAC), so the percentage of children and women screened is an estimate of access to the TSF.

Table 3.26 shows that about 50% of *children* participated in the most recent screening in tranche 2, and 30% in tranche 3; these percentages did not change significantly between baseline and midline. The reason most often reported for non-attendance changed from "other reason" at baseline in both tranches, to "did not know about date/time" at midline in both tranches. About 15% of the children screened then received supplementary food (see table 3.26, second row). Of the entire child population at midline, 11% in tranche 2 and 7% in tranche 3 received any supplementary food from the most recent TSF screening.

As for *mothers* attending screening while pregnant (table 3.27) about 25% of these were screened in tranche 2, and just above 10% in tranche 3 – relative proportions similar to child screening – and this too did not change between surveys. Of those screened, about half received supplementary food while pregnant, although this fell in tranche 3. Thus roughly 1-10% of pregnant women received supplementary food (e.g in tranche 2 approximately 0.25 participating * 0.50 of these getting food); however this fell substantially between base- and mid-line in tranche 3.

Indicators of TSF participation/receipt are needed to control for this program as a possible confounder; and along the way (if feasible) to assess its nutritional impact. Although the receipt of TSF is estimated as fairly small – less than 11% for children – there may also be community level effects when villages participate. The estimate was derived by categorizing EAs by those in which children received supplementary food from the most recent screening or not (see Methods section for further description), with 57% of EAs in tranche 2 and 38% in tranche 3 receiving any supplementary food from the most recent TSF screening.

E. Changes in stunting ascribed to CBN and TSF, controlling for potential confounders

Changes in mean height-for-age and stunting prevalence were significantly greater in EAs with higher levels of CBN program implementation, as shown in section 3 B (see also table 3.15). We now need to examine whether these changes might have been due to other programs taking place in the tranche 2 and 3 woredas, and/or whether they can be ascribed instead to changes in socioeconomic and environmental factors (i.e potential confounders). TSF is the primary other nutrition program that is relevant, and this is examined first. Then the other confounders are considered.

i. Distinguishing CBN and TSF impacts on stunting

The association of TSF and CBN participation at EA level with changes in mean height-for-age and stunting is shown in table 3.28 and figure 3.8. First, there is no association in the locations of the two programs – seen by the n's in the table: EAs with TSF are no more likely to have CBN, and vice versa (chi-square tests on the distribution of cases are quite insignificant).

The results for tranche 2 indicate that only those EAs participating in either TSF, or CBN, or *both* showed significant reductions in stunting prevalences (with similar results using HAZ). Reasonably similar results were found in tranche 3. However a minority of EAs had no programs (n=11), and this should be taken into account interpreting these results.

Higher prevalence EAs were targeted for CBN and or TSF, as can be seen in figure 3.8. Here, those with either program, or both, improved, compared to the no-program group. Either program appeared effective (significant on regression controlling for each other), and combined effects were less than additive. The results for tranche 3 showed no association with TSF, but an association with CBN.

EAs with higher starting prevalences that received either or both programs improved to a prevalence the same, or *better,* than without-program. The size of the effect of these programs, over the periods studied, was around 10-15 ppts lowering of stunting prevalence. Tranche 3 showed less consistent but broadly similar results. As will be seen, these estimates persisted when controlling for other potential confounders.

ii. Controlling for SES and environmental factors

Regression analysis (OLS) was used for examining potential confounders. The variables of interest here (CBN and TSF) together with all the relevant SES variables were regressed with changes in HAZ or stunting as the dependent variables. A drawback is that the SES variables are collinear, so we do not see their individual effects; moreover this risks over-controlling so no variables emerge significant. The models are shown for HAZ and stunting in table 3.29, and these are focused on here.

For tranche 2, the regression coefficient for CBN implementation, controlling for all other factors, with change in HAZ as the dependent variable, was 0.377 (p=0.151) and 0.580 (p=0.005) in tranche 3; regression coefficients with change in stunting as the dependent variable were -

0.112 (0.070) in tranche 2 and -0.138 (p=0.038) in tranche 3. These coefficients remain largely unchanged (in both size and significance) from the CBN implementation coefficients without controlling for potential confounders (change in HAZ as dependent variable in tranche 2 was 0.377 [p=0.113] and 0.577 [p=0.003] in tranche 3; change in stunting prevalence as dependent variable in tranche 2 -0.120 [p=0.020] and -0.124 [p=0.046] in tranche 3). These coefficients are taken as evidence that the observed changes of HAZ and stunting may plausibly be ascribed to the CBN program, with participation as defined in the dummy variables.

Similar calculations can be done for TSF, also shown in table 3.29.

WAZ is associated with CBN implementation, in the full model used for stunting (as seen in table 3.29), coefficient for tranche 2 was 0.328 (p=0.072) and tranche 3 0.295 (p=0.029). This is in the same direction and of similar size as stunting. Wasting was not associated with program activity.

Of the independent variables, when all together in the model none come through as strongly significant, probably due in part to the collinearity. Substituting the level of each independent variable for changes in that variable did not alter the substance of the results quoted here, results not shown for brevity.

Interactions of program variables with selected independent variables would be of interest for future research, to investigate whether the impact of these programs varies with their level – for example do children of educated caregivers improve more?

iii. Taking account of food insecurity, drought, and seasonality

Food insecurity. The level of household food security at baseline and midline can be assessed from the 'Household Hunger Scale' (HHS), which results from 3 questions concerning food access as described in the Methods section. Results are shown in table 3.30. The calculated prevalence of 'little or no household hunger' improved by 7 ppts from 81% to 88% (p<0.01) in tranche 2 between base- and mid-line. When the change in this indicator is included in the regression models discussed above, it causes little change to the program coefficients. A slight decrease in 'little to no household hunger' is seen in tranche 3, dropping from 86% to 83%, but the change is not significant.

While at least by this direct measure of household food security there is no evidence that food availability is accounting of any outcome changes, we need to also consider seasonality and drought, for which we do not have direct measures for these samples.

Drought. Overall, drought affected parts of Ethiopia worse in 2011 than in 2009-10, but there was some drought in 2010 (see table 3.32). For example, FAO gives estimates of cereal production (million MTs) as 2009, 16.8; 2010, 17.4; 2011, 15.5 (FAO, December 2011, p 14). Woredas assessed as most affected and needy are identified as 'Hotspots' and targeted for assistance, for TSF and other programs. Hot Spot Priority woredas are selected several times per year; the selections prior to and following the baseline and midline surveys for both tranches were investigated to determine the percent of EAs categorized as Priority One woredas (the most severe food insecurity) out of the EAs surveyed. These percentages are then compared between baseline and midline to determine if the EAs surveyed were more often classified as food insecure at either point in time.

As seen in table 3.31, in tranche 2 (for instance) 34 of the 60 woredas with information, i.e. 57%, were declared as Priority One Hotspots, around the time of the baseline survey (in April 2009), compared to the national average of about 30% at that time: tranche 2 was more drought affected than the average at baseline (i.e. a higher percentage of hot spot woredas). By midline tranche 2 was nearer the average (22%) at 33%. Tranche 3 started nearer the average, and did not change much by midline.

Seasonality. Ideally, baseline and midline surveys are conducted at the same time of year to avoid seasonal effects on malnutrition. However, due to logistics, the baseline and midline surveys for both tranches 2 and 3 were conducted during different seasons. As shown in table 3.32, the tranche 2 baseline survey was undertaken in June and July of 2009, primarily a hunger season in Amhara, Oromia, and Tigray, while edging into the harvest season in SNNPR. The baseline for tranche 3 was conducted in March and April of 2010, during the short rainy season in all regions but SNNPR; this is usually also a time of shortage, earlier in the year. Both midline surveys were conducted in September and October of 2011, primarily an end of hunger and/or harvest season in all regions.

As to the size of this seasonal effect, for comparison, seasonal effects on wasting in the Horn of Africa were estimated as roughly up to 2-3 ppts, (Mason et al, 2010).

Combined effects. A summary of the probable combined effects is given in tables 3.32 and 3.33. Without-program, we would probably expect deterioration in tranche 2, and more so in tranche 3. This is what is seen in the no program group in figure 3.8 for tranche 2, where there is no significant change in that group. For tranche 3, the without program trend is also not significantly different from zero.

4. DISCUSSION

The Community-based Nutrition (CBN) program was conceived and designed in 2007-8, and formally launched in mid-2008. Specific objectives were set for anthropometric and IYCF indicators, which were suitable for estimating with repeated surveys, as anticipated in the planning documents (FMOH, 2008 & World Bank, 2008). The evaluation used repeated surveys at EA level, plus routine data from the weighing programs, see figure 1.1 in Background. No predetermined comparison groups were possible. However, during the analysis it became clear that there was considerable variation in implementation, thus *de facto* internal comparison groups were feasible:

- For tranche 2, there was substantial variation in the time VCHWs spent on program activities (see figure 3.2), thus allowing two groups to be defined, with less or greater than mean VCHW activity levels;
- For tranche 3 the definition of a *de facto* internal comparison group was more straightforward, as 17 of the 61 EAs were considered not to have started the program activities at all, as these were in woredas that had not begun reporting routine data by the midline survey.

These comparisons, and comparing with the long term trends from all available DHS data[9] , showed that:

- The changes in stunting within tranche were statistically significant;
- The improvements in stunting were significantly greater than the secular trend (2000-2011), and;
- The changes in stunting showed significant impact:
    - In tranche 2 the group with higher VCHW activity improved stunting significantly faster than the low activity group;
    - In tranche 3,the reporting EAs, compared with the non-reporting (17 EAs) improved stunting significantly faster;
    - These changes remain significant when controlling for potential confounders.

These findings are central to the evaluation and discussed in more detail below (see table 4.3).

A. Why stunting?

The most consistent trends were in stunting means (HAZ) and prevalences. As expected, wasting fluctuated much more between surveys: these fluctuations through time (within area or livelihood group) stem mainly from season and drought (Chotard et al, 2010). Thus underweight (which is largely due to stunting but in part to wasting) fluctuated more than stunting.

Stunting is expected to be the most stable indicator, but only measures linear growth. Underweight is in principle the best, measuring both soft tissue (muscle, fat, organs, brain, etc) and linear growth (ACC/SCN, 1990), and capturing changes in both wasting and stunting (with which it is highly correlated). Stunting, because of its stability, being less affected by seasonal changes and short-term drought, was expected to be the indicator likeliest to show significant response to intervention, and was the first looked at. Stunting has become the anthropometric

[9] Similar trends were seen in the Welfare Monitoring Surveys (WMS) (as seen in the 'Niphorn' report by Chotard et al, 2007, accessed at http://www.tulane.edu/~internut/Trial/RSRC.htm

indicator of choice: e.g. 'I am talking about stunting – one of the greatest human inequities and social injustices of our time'[10]. Certainly here it turns out to be the most stable and significant indicator.

B. Length of time of program implementation

In interpreting these results, it is important to keep in mind the varying lengths of time of program implementation. Figure 3.1 shows this; note that this is calculated from the reporting times from the routine data, found to be a better source. While we have taken the time of implementation for tranche 2 as 2.3 years – which is the total time since launch -- we should note that the median implementation is around 21 months, due to delays in launching in some woredas. This consideration is more important in tranche 3, where 17 EAs were estimated not to have started by the mid-term, 18 months after program launch. The median time of implementation is actually 6 months, however for calculations here the time-since-launch of 1.5 years is used for consistency.

Programmatically, while it is encouraging that in tranche 2 all targeted woredas were reporting by month 15, the delays exemplified by tranche 3 will lower the impact because of non-performance; and may also lower the quality as knowledge and skills from training and orientation fade with time.

C. Process results

The CBN was implemented tranche by tranche (show in table 1.2). The first estimates of coverage came from the CBN data in 2010, tranche 1, suggesting that about 30-40% of the under-2 child population in the implementing woredas were being weighed (Hoblitt & Mason, 2010): this calculation could not be precise since there was no indication of how many times each child appeared in the weighing data. Subsequent routine data confirm this estimate (Buback & Mason, 2011; White and Mason, 2011). In the 2011 data, a total of nearly 33,000 VCHWs were reporting from the tranche 2 woredas (table 1.3), and VCHW:child ratios were estimated as averaging 1:11.

A number of questions in the CBN evaluation surveys, in the household and cluster level questionnaires, addressed different aspects of coverage, participation, and intensity of program activities. Training was extensive, for example with about 80% of VCHWs having received refresher training (see table 3.2). Participation in some aspects of the health and nutrition programs was around 70% of the child population (and caregiver), in terms of at least one contact with HEW or VCHW (tables 3.6 and 3.7).

These and other results can be compared with program objectives and coverage indicators specified in the PIM (FMOH, 2008, p 110), as discussed below.

[10] From a Keynote Speech given by UNICEF Executive Director, Anthony Lake on April 1st, 2012 in Kampala Uganda. Available at: http://www.ipu.org/conf-e/126/unicef.pdf

Most of the process indicators (see 'Aims' in table 4.1), for which estimates could be made from the survey data, were in line with or exceeded targets. For guidance, we have scored these for the extent to which they appear to reach aims, taking account of the planning documents (frequently giving only year 3 and year 5 targets); these scores are given in the last column.

For example, the percent of children weighed (first indicator in table 4.1) was given a target of 50% by year 3, with no baseline estimate. This presumably assumed that all weighing comes from the CBN, although in fact some was already occurring. The increases in percent weighed, tranche 2 in year 2, was from about 10% to 30-33%, in line with the aims. The increasing trend is of 20 ppts per year.

The training and refresher training of HEWs and VCHWs are the next two indicators in table 4.1. Training of HEWs is scored as satisfactory, comparing the year 1 and 2 levels for tranches 2 and 3 (43-55%) with the year 3 aim of 60%. Refresher training clearly well exceeded planned rates, at around 80% compared with a year 3 aim of 40%.

The VCHW ratio to population was proposed as 1:50 households. The intended target group is 0-2 yr old children, since there is approximately one under-5 child per household, 0-2s should be about two-fifths, or 40%, thus the planned ratio is estimated at 1:20. As can be seen from table 4.1 this ratio is about on target. This is from the survey data, and we should note that we estimated 1:11 from the CBN routine data; the survey data should be the more accurate.

The percent of pregnant women receiving iron/folate, which was part of the micronutrient section of the survey questionnaire, matched one of the specified results indicators, and may also reflect more general use of health services. In any event it is included here, and shows a substantial increase, providing more evidence for the effective outreach of the health services and HEW and VCHW activities.

While the objectives stated may often have only a limited basis for their specification, the comparisons taken together show rather clearly that the project was implemented at a level that usually met or exceeded the objectives' process indicators. This is an achievement, both in implementation, and indeed in monitoring – as a result of which we know quite a lot about how the project has been proceeding.

With other data presented in the Results section, we can conclude along the following lines:

- Coverage by woreda reached 100% in tranche 2 and 70% in tranche 3. Implementation by woreda could progress faster after the launch of each tranche – 28% of tranche 3 had not yet started after 2 years – thus follow-up from the regional level to the woreda level to ensure more rapid implementation, fixing whatever is causing the delay, should be another priority;
- Participation was probably at about one-third of the children in active woredas being substantially involved in the CBN, e.g. were weighed. However, two thirds are less engaged or not at all: raising this participation rate, in kebeles where the program is already active, should be a priority going forward;
- Probably another third of the under-3 children had some influence from the program – e.g. as measured by contact with the VCHW, and this contact increased to over 80% (at

least one contact) and at least 40% with regular contact, likely providing informal inputs to behavior change;

- The time spent on CBN activities by the VCHW averages 7 hours each week, with a considerable spread (see figure 3.2) and much variation in the different aspects of the CBN activities that time is spent on: some consideration of providing guidance to optimize the VCHW's time allocation could be useful. (Further analysis of the existing data could investigate this some more.)
- The ratio of VCHWs to children was right around 1:20, and this is within the range of program intensity consistent with a significant impact on child nutrition.
- Overall the program was successfully implemented to the level foreseen in the planning documents, or better.
- In view of this, we should expect to be able to detect changes in IYCF and other goals of the CBN (and the HEP); and then on child nutritional status.

D. Impact on infant and young child feeding (IYCF)

The results given in table 3.10 have shown clearly some important and significant changes in IYCF. Some of these indicators are extracted, compared with program aims in table 4.2, and again scored for their indication of success.

Exclusive breastfeeding under 6 months, dietary diversity between 6 and 23 months, and the percent getting a minimal acceptable diet at these ages, all improved substantially. In some cases, like the first on breastfeeding, the extent of exclusive breastfeeding among 0-6 month children is already high – around 80% – and remaining so. This suggests for program planning that the advice from HEWs and VCHWs should be directed towards encouraging continuation of an already beneficial behavior; and for everyone to be on guard to prevent threats (e.g. from the infant formula marketers) to this excellent practice.

On the other hand, the advice to start breastfeeding within one hour of birth has not been effective – if given – in increasing this practice, although it is already at a level of 50-60%. The aim was 83% in year 3, from a baseline estimated (presumably from survey results) at 69%; and the level estimated from the evaluation surveys was a steady 50-60%. Training on this issue should be reviewed.

Complementary feeding has improved in terms of diversity of diet rather than timing of introduction of complementary foods – the indicator (the third in table 4.2) exceeds aims at 50-70%, but was not increased significantly by the CBN. On the other hand, the indicator minimum acceptable diet showed a major improvement (fourth indicator, table 4.2), although this was not included in the PIM indicator series (FMOH, 2008, p 110-116).

The counseling to continue feeding children with diarrhea seems to have been highly effective, increasing from less than 10% to 25-38% (see table 3.11). A similar success was seen (shown in table 3.23) in a decrease in the habit of eating less during pregnancy ('eating down'), based on fears of difficult delivery of a larger baby. This is included in the PIM indicators, but with no quantitative goals given.

Overall, the improvements to child feeding practices seem likely to contribute to better growth and nutritional status in children. Two further analyses remain to be done:

- First to check whether the changes in feeding practices are confounded by, for example, improving SES as proxied by house quality, e.g. roofing; this would be the same as described later for anthropometry; time has not allowed for this, but experience with the data suggests that such potential confounding is not having a significant effect on the outcomes; still, this should be investigated..
- Second, to explore whether better feeding practices are linked to better growth in these datasets: this is a significant piece of research which has not yet been done.

E. Impact on stunting

Significant overall improvements were found in stunting by tranche between the baseline and midline surveys. This was important: if no differences had been found – even if this was due to a countervailing worsening without-program trend – it would have been difficult to draw many conclusions.

The variations discussed earlier in program implementation provided for *de facto* internal comparison groups. Specification of these groups likely involves some degree of self-selection, or selection bias. Thus the differences found are re-estimated controlling for potential differences between participants and non-participants; or, put otherwise, potential confounders. This was done in the regressions shown in table 3.29, showing that the coefficients for with/without remained at a similar size and significance controlling for a variety of potential confounders.

In evaluating differences in stunting related to CBN activities, we first compared program-related trends with norms – i.e. the underlying trend as determined in national surveys, by DHS, estimated as -1.2 ppts/year (see table 3.12). For tranche 3, the 'reporting' group was used for this comparison. For tranche 2, all the EAs were included. Results are summarized in table 4.3. Here the change in stunting for the with-program groups, of -15.7 ppts and -8.0 ppts in tranches 2 and 3 were significantly higher than that predicted for the same period of 1.2 ppts/yr.

The differences in changes in stunting between the high/low activity, and the reporting/non-reporting groups, were also significant (i.e. in tranche 2, between -15.7 ppts and -2.8 ppts, p= 0.02 in tranche 3, between -8.0 ppts and+ 4.4 pts, p= 0.05). The reasons for the higher starting prevalences in the high activity/reporting groups are not known except that these programs are targeted with priority given to poorer or more food insecure woredas. However the data did not allow us to compare with/without program at different starting prevalence levels.

We conclude that the with-program improving trend is considerably higher than expected without the program. For tranche 2, the size of this effect, adjusted for potential confounders (table 3.15), was 11.2 ppts improvement over the 2.3 year period attributed to the program. This is equivalent to the program adding 4.9 ppts/year improvement, over and above the underlying trend. For tranche 3 the effect size of 13.8 ppts, similarly calculated from table 3.15, is 9.2 ppts/year. The difference between tranches may well reflect the observation from the CBN routine data, as shown in figure 1.1, that there is more rapid improvement in the first year, which then slows down to a lower improving rate in subsequent years. This has been observed in a number of other programs (WHO, 2012). However, the reasons are not established, although they can be hypothesized (e.g. worst off children benefit first; improved access to health services deals with immediate needs, such as deworming; etc).

The change in the anthropometric indicators can be compared with the program aim, as found in the PIM, this is shown in table 4.4. Stunting indicators greatly exceeded targets.[11] The target set was clearly less ambitious than it might have been. Wasting does not seem to be related to the program, and most experience is that this only responds to short-term factors, such as diarrhea or other intermittent illness or acute food insecurity. Underweight does about meet the target, but is probably less satisfactory in this case because the fluctuating wasting increases the variation in underweight.

Disentangling effects of TSF and CBN appeared to show that both had an effect, but this was not additive: participation in either program, or both, was associated with reducing prevalences of stunting.

In sum, the evidence is plausible that the CBN activities brought about a substantial decrease in stunting in children of under-3 years of age. The important features contributing to this impact are likely to be:

- A high ratio of VCHWs to children, estimated here as somewhere between 1:18 and 1:19: this is in the range where substantial impact is expected (Mason et al, 2006 & WHO, 2012)
- Regular contacts between mothers and the VCHWs and HEWs.
- Effective counseling of mothers on feeding and caring practices.
- Referral of sick children for medical treatment and if necessary supplementary feeding.
- The VCHW is in and of the community, and thus communicates with mothers regularly on an informal basis.

More evidence on the relevant impact of different activities might well be obtained from further analysis of the evaluation datasets. However, given limited time and resources – particularly for advanced analysis – and the limitations of the present data, especially lack of pre-determined comparison groups, it may be that the best plan is to use these results for current decision-making, and move ahead to more solid evaluation in future surveys. As we understand it, the baseline survey plans for upcoming tranche 5 of CBN include the possibility of pre-determined comparison groups.

[11] Note that the original stunting targets in the PIM (46% to 40%) are quoted in NCHS standards while CBN evaluation results are quoted in WHO Standards. If converted to WHO standards, the PIM targets change to 52% to 46% (cf. 46-40%). Even after the conversion, the CBN evaluation results still exceed this target, as the change seen is greater than the 6 ppt change aimed at by the PIM (whether in NCHS or WHO standards).

F. Examples of impact on performance indicators

In important aspects, the project performance exceeded expectations. Four examples chosen to demonstrate this are given in figure 4.1. As described extensively above, stunting rates were improving several times faster than the aim. Project implementation can be assessed by numbers of children weighed, and this percentage grew at the rate specified in the PIM. As to indicators of IYCF: exclusive breastfeeding has maintained far higher levels than the aim specified in the PIM. As a good example of caring practices the proportion of children given the same or more food during diarrhea, increased at or better than the rate in the PIM target. These examples are chosen from the range available to illustrate in summary form the success of the program.

G. Recommended program decisions

The CBN program appears worth continuing, and expanding in area coverage. The data for the first 2-3 years, from evaluation surveys plus CBN routine data, showed that the program exceeded the stated aims, both in terms of implementation and outcomes. Indicators of infant and child feeding practices, and nutritional status objectives, were met or exceeded. More rapid implementation in woredas after the program launch is indicated to increase coverage.

The participation rate in areas where the program is implemented reaches probably 30-40% of the 0-3 year old children. Increasing this participation provides perhaps the greatest potential for increased impact. However some operational research may be necessary to elucidate reasons for the limited participation, and hence ways to increase it.

Attention is needed to initial and refresher training and support to VCHWs. While this was high in the evaluations, this needs to be sustained to maintain incentives.

However, extensive program revision under way, with the 'Health Development Army' replacing VCHWs. As we understand it, some nutrition activities will continue to be undertaken by the frontline workers, although the training in nutrition activities may be much reduced (to half a day, it has been suggested). Clearly it is a policy decision as to how far to continue the nutrition activities, which it can now be argued have demonstrated success. Adequate training and monitoring will be crucial as the shift to the Health Development Army proceeds.

H. Recommended policy decisions

Policy decisions refer to whether and how to initiate and sustain programs aimed to improve child nutrition. This can apply within Ethiopia, and to policy decisions on whether and how to run similar programs elsewhere.

A first consideration is why to have such programs – why are they advocated for use of scarce resources. The case is well made in the GoE's own words, from the project documentation (FMOH, 2008, p4):

'The consequences of malnutrition for Ethiopia if no action is taken are enormous. The greatest functional consequences of malnutrition for children are illness, and death; and for those who survive, mental impairment and reduced capacity to produce and contribute to the economy of the country. These consequences of malnutrition are often not fully appreciated because they are hidden...'

Plus the recognition that:

'Malnutrition is one of the main health problems facing children and women in Ethiopia. The country has the second highest rate of malnutrition in Sub-Saharan Africa (SSA).'

Tackling malnutrition is not suggested as necessarily the top or only priority. Sustained support to the HEP is clearly vital (and without it the CBN would not be possible anyway). Equally, a shift towards more multi-purpose frontline health workers, with a very high ratio to households, may open up new opportunities. The conclusions of a recent review of the potential contribution of community health workers stressed, however, that clearly defined roles and a limited series of specific tasks are likely to lead to better performance (Haines et al, 2007).

The developments in health in Ethiopia can hopefully continue to give priority to maintaining the nutrition counseling, weighing, and other activities of the CBN at the current levels. This could be one role of a multi-purpose frontline health worker, or alternatively there could be some specialization of these to include nutrition. The best way forward will need careful consideration. Either way, adequate time for training and retraining, and supervisory support, will be needed to maintain the impact. We also recommend that data systems be further developed to keep track of the effects of the changes.

In what contexts should we recommend such programs elsewhere (with a focus on Africa)?

Some considerations are as follows:

- An organization is in place – often within the health sector, similar in principle to the Health Extension Program in Ethiopia – upon which a community-based nutrition program can be built;
- The causes of malnutrition can be influenced by behavioral change and other community activities under the VCHW (for instance, if malnutrition is primarily due to an acute food crisis, or an epidemic of infectious disease, then CBN would not be the first priority)
- Institutions and/or agencies have the capacity to provide guidance, resources, and training (and refresher training), and other incentives at local level, to support the CBN program
- The capacity exists or can be built for adequate monitoring and evaluation, and this can be provided with sufficient resources
- The government gives priority to improving nutrition: in policy terms, assigning resources, personnel (including job descriptions, career development, etc), and building institutional capacity.

The results reported here may be seen in the context of a number of other evaluations getting similar results. They can also be compared with some showing negative results; the lessons have contributed to the considerations laid out above. In Africa, evaluations of large scale programs in Madagascar (Galasso & Yua, 2006; Galasso & Umpathi, 2007) and Senegal (World Bank, 2001; Alderman et al, 2009) have shown similar rates of improvement (figure 1.2). These rates are measured in underweight rather than stunting, but these two are closely associated. In Tanzania, the Iringa project in the 1980's showed a very similar pattern to that seen here, especially in the CBN routine data which covers a longer period: in Iringa, the underweight prevalence fell from 55% to 40% in the first 1.5 years (10 ppts/yr), then slowed to around 1.5 ppts/year improvement (Mason, 1996). Much the same pattern is seen in figure 1.1.

I. Recommendations on evaluations

The evaluation of the CBN program is planned to continue, and these recommendations begin by applying to this ongoing process. However, there are some lessons for other evaluations which will be suggested.

First, it is important to take opportunities to include comparison groups in future evaluation designs. Tranche 5 of the CBN is about to be launched, and it is understood that now the planning is such that some woredas will receive no outside support for CBN implementation. These can act as a comparison group, and should be surveyed at the same time as the externally supported woredas in tranche 5. Data should be analyzed after this baseline, to check how comparable the comparison group is, in terms of nutritional status, other programs, and contextual factors.

Second, we so far only have the midline data, after 1-2 years of implementation. Assessing the longer term trend is also important, particularly as there is evidence, from the CBN routine data and results from elsewhere, that the rate of improvement slows after the first 1-2 years. A five year period between base and end-line was planned, although there could be some flexibility in this. The endline survey certainly for tranche 2, better still also with tranche 3, needs to be planned for. Conceivably this could be coordinated with other surveys – perhaps the midline for tranche 5.

Third, it is crucial that the data be comparable across all the evaluation surveys. The household questionnaire should remain the same for the key questions and variables, such as those used in this report. However there a considerable number of questions and associated variables that are not being used, and these could probably be cut out of future questionnaires. The cluster level questionnaire has proved crucial and must continue; in fact there could be some additions to this.

Fourth, a key issue has remained unresolved: developing the national capacity for these types of analyses, and indeed for survey design and implementation. Until this is seriously and realistically addressed, analysis and interpretation will continue to depend on international assistance.

For other evaluations, some aspects of the present design – hopefully expanded to include pre-determined comparison groups – may prove useful. Using repeated measures at the cluster level, by using the same clusters for base-, mid-, and endline surveys and re-sampling households/children, adds statistical power, and the findings here would not have been possible without this design feature. This is different to many evaluations, which use repeated cross-sectional surveys with completely new sampling each time. Using the same clusters is no more difficult, and in some ways easier.

Finally, more impact evaluations are needed. These do not need to be randomized controlled trials – in fact they cannot be if they aim to assess the effectiveness of large scale operational programs. They need to be carefully designed to include some form of comparison; sometimes, like here, these can be specified *post facto*, based on sufficient variation in program implementation to give internal comparisons. Moreover, not every program needs an impact evaluation. Some should proceed with monitoring process variables, primarily through administrative routes, to ensure the adequacy of the program implementation. It would be better to adequately fund a limited number of selected programs (and analyze the results) than try to evaluate impact too widely. We can note too that impact evaluations are often only available after programs have finished, so their results are useful for decisions about other future programs, but not for the program being evaluated. Thus they may legitimately expect funding from sources other than the program being evaluated.

J. Data limitations and advantages

The inability to utilize external comparison groups in the survey design made it more difficult to attribute impact to the CBN program. Additionally, it was not possible to randomly assign program implementation. However, the survey was designed to use repeat measures at the EA level, with each acting as its own control. While this is analytically more powerful, it resulted in a small sample size, as the EA, rather than the individual, is the unit of analysis. Five clusters in both tranches were missed during the midline survey, resulting in a smaller sample size.

The two baseline surveys were conducted by a different firm than the two midline surveys. As a result, there were substantial differences in database structure as well as actual differences in the structure of the questions and possible categorical responses in the questionnaires. Thus, several variables could not be analyzed as the data collected at baseline and midline could not be aligned. Considerable time was needed to reconcile variables in baseline and midline datasets into similar formats for aggregation and analysis.

Measurement of anthropometry also slightly differed between surveys, with for example the two baseline surveys showing a large number of cases in which the child was measured for length incorrectly according to their age (e.g. children under-2 measured standing instead of lying). On the other hand, in the midline surveys there were no cases in which the child was incorrectly measured lying/standing.

Substantial age heaping was found in all four surveys, particularly at 12 and 24 months (as described in Annex). Details are in the Annex. Extensive age re-calculation was needed and exploration of new z-score variables calculated using non-exact ages was necessary to determine if heaping could be further reduce, although exact and non-exact age calculations resulted in very similar outcomes.

While the survey questionnaire captured large amounts of useful information, a number of important factors were omitted. Examples are: income; women's health or age at first birth;

therapeutic feeding for severely malnourished children; and access to and contribution of other programs e.g. the Productive Safety Net program (PSNP).

A particular design feature was repeated measures at the cluster level, which proved crucial in getting enough statistical power. The staggered implementation design was appropriate, of course better with comparison groups. The EA questionnaire was also crucial and must be continued, with additions related to other program participation, village level factors (such as roads, access to services). Continued analysis and triangulation with the CBN routine data is needed.

As discussed earlier, the most important design development will be to have external comparison groups. However, the questionnaires, while needing some streamlining in view of the considerable number of variables never used, should remain otherwise the same for comparative purposes.

TABLES AND FIGURES

Table 1.1: National Nutrition Program components

Component 1: Strengthening Service Delivery
Sub-components:
A: Sustaining EOS/ TSF and Transitioning of EOS into HEP
B: Health Facility Nutrition Services
C: Community Based Nutrition
D: Micronutrient Interventions
Component 2: Strengthening of Institutions for Nutrition Policy and Program Implementation
Sub-components:
A: Strengthening Human Resources and Capacity Building
B: Advocacy, Social Mobilization and Program Communication
C: Strengthening Nutrition Information/Surveillance, Monitoring and Evaluation, and Operations Research
D: Strengthening Multi-sectoral Nutrition Linkages

Source: (FMOH, 2008)

Table 1.2: Descriptives of CBN implementation by tranche

	No. of woredas	Timing of implementation
Tranche 1	39	July 2008 – January 2009
Tranche 2	54	July 2009 – October 2010
Tranche 3	77	August 2010 – Present
Tranche 4	58	August 2011 - Present
Total	228	-

Figure 1.1: Underweight prevalence (NCHS standards, <-2 SDs) in children under-2 weighed from CBN monthly routine data, August 2008 to March 2012, by region, in *tranche 1*

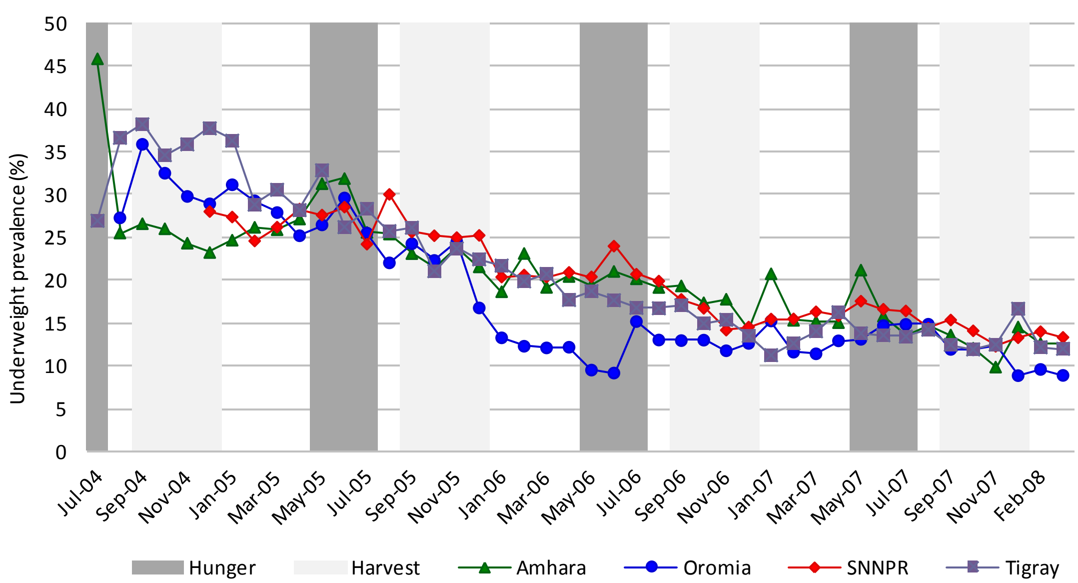


Table 1.3: Number of VCHWs trained (highest accumulated number reported), reporting, and VCHW intensity between January and June 2011 in Tranche 2, from CBN routine data

	No. VCHWs trained	No. VCHWs reporting	VCHW intensity [a] *Jan-June 2011* [b]
Region			
SNNPR	5429	4370	1:14
Oromia	4921	3217	1:15
Amhara	10955	7749	1:9
Tigray	13621	10043	1:7
All four regions	34926	32966	1:11

[a] VCHWs to children

[b] Data averaged between January and June of 2011

Source: White & Mason, 2011

Figure 1.2: Population sustained rate of underweight reduction (ppts/yr) compared to program intensity estimated as CHNWs per 1000children, as part time equivalents (0.1 FTEs) (WHO, 2012)

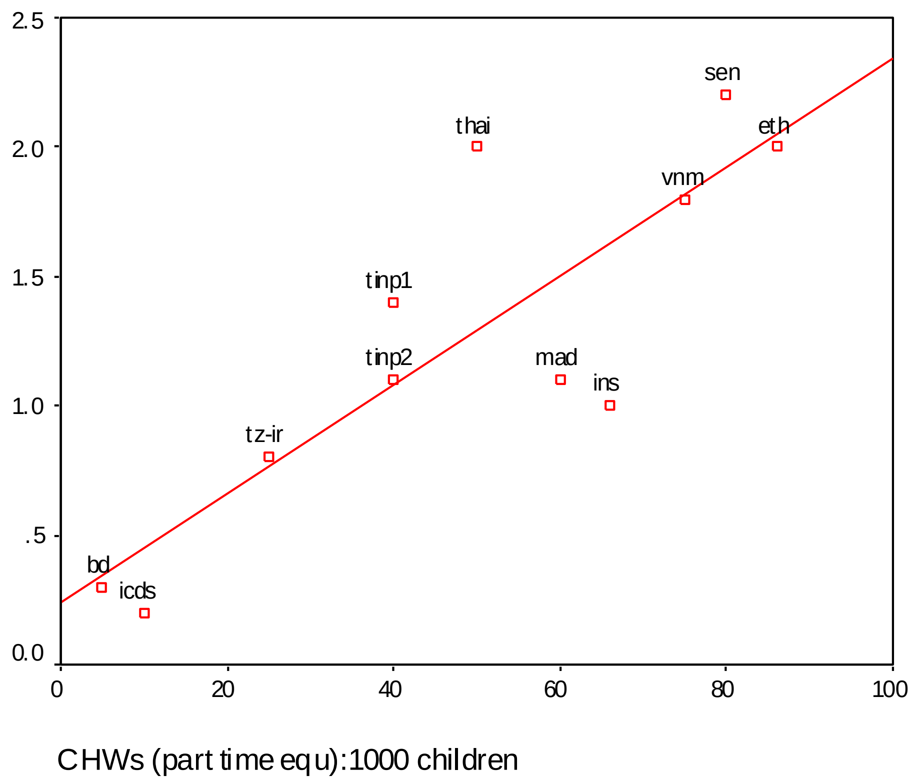


Figure 2.1: Proposed phasing for National Nutrition Program Evaluation

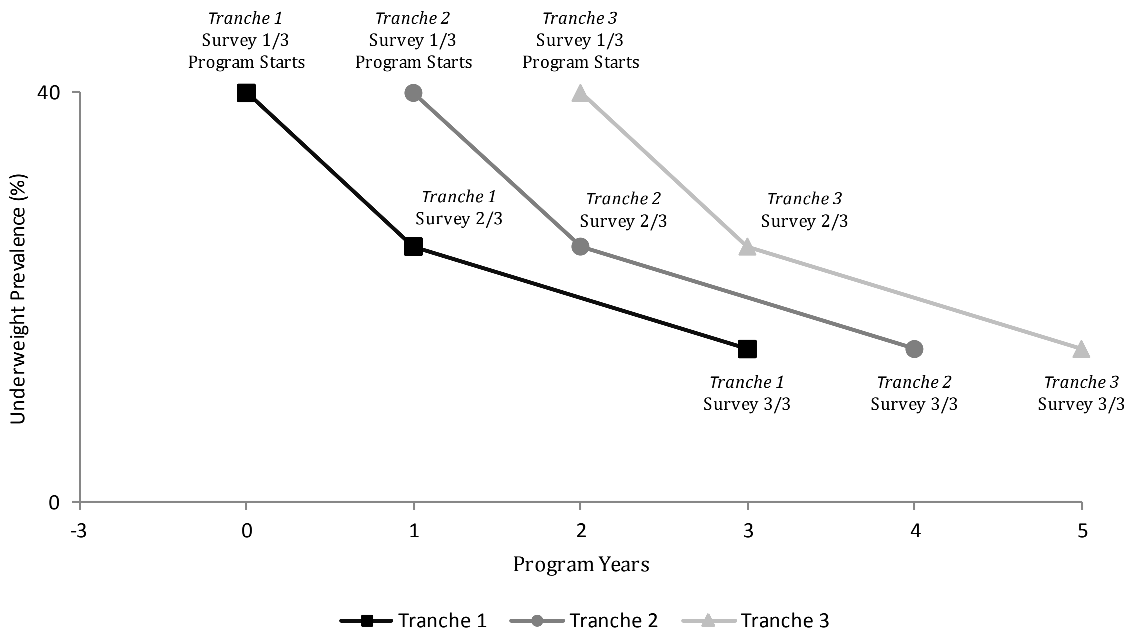


Table 2.1: Definitions of key indicators

Key Indicators	Definition	Source
<u>Program participation</u>		
Tranche 2		
Intensity of CBN activity	Number of reported hours VCHWs spend each week on house-to-house visits, counseling, teach, and community discussion, dichotomized into EAs with high (>7 hours per week) and low (<7 hours per week) intensity.	EA level data collected from HEWs during evaluation
Tranche 3		
Length of program activity	Number of months between first report of CBN routine weighing data and midline survey (October 2011)	Routine data
<u>Outcome variables</u>		
Anthropometry		
Z-score	Difference in height-for-age, weight-for-age, and weight-for-age mean values, created by subtracting the midline aggregated estimate from the baseline aggregated estimate	Individual level datasets
Prevalence	Difference in stunting, underweight, and wasting prevalence values, created by subtracting the midline aggregated estimate from the baseline aggregated estimate	Individual level datasets
IYCF	All indicators created using WHO guidelines (WHO, 2010)	Individual level datasets
<u>Contextual variables</u>		
Socioeconomic status		
Toilet facility	Improvement in toilet facility, created by categorizing EAs into those that experienced any improvement in use of 'improved' toilet facilities and those that experienced no change or deterioration between baseline and midline	Individual level datasets
Drinking water supply	Improvement in drinking water source, created by categorizing EAs into those that experienced any improvement in use of 'improved' drinking water source and those that experienced no change or deterioration between baseline and midline	Individual level datasets
Caregiver education	Improvement in caregiver education, created by categorizing EAs into those that experienced any improvement in caregivers level of education and those that experienced no change or deterioration between baseline and midline	Individual level datasets
Roofing material	Improvement in roofing material, created by categorizing EAs into those that experienced any improvement in use of 'improved' roof materials and those that experienced no change or deterioration between baseline and midline	Individual level datasets
Health Extension Program		
ANC attendance	Improvement in attendance of ANC, created by categorizing EAs into those that experienced any improvement in utilization of any ANC and those that experienced no change or deterioration between baseline and midline	Individual level datasets
Other programs		
TSF participation	Receipt of supplementary from TSF, created by categorizing EAs into those that had any children receive supplementary food during the last TSF screening and those that had none	Individual level datasets

Figure 3.1: Frequency of length of program activity in months, by tranche

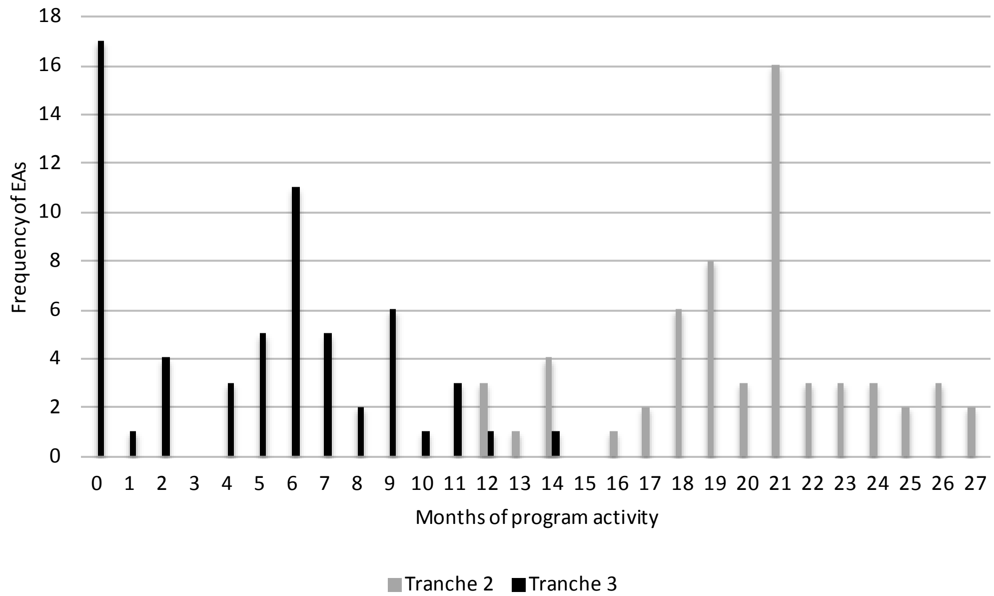


Table 3.1: Descriptives on length of program activity (number of months reporting CBN routine data prior to midline survey), by tranche

	Tranche 2	Tranche 3
Descriptives		
Number of EAs included in baseline and midline	60	61
Start month range of CBN reporting	Jul 2009 – Oct 2010	Aug 2010 – Sept 2011
EAs with >0 months CBN reporting [a]	60 (100%)	44 (72%)
EAs with 0 months CBN reporting [a]	0 (0%)	17 (28%)
Range of length of program activity	12-17	1-14

[a] As of October 2011

Table 3.2: Descriptives on HEW and VCHW training at midline (%)

	Tranche 2	Tranche 3
HEWs trained in CBN		
% of EAs with 100% of HEWs trained	91.4	89.5
% of EAs where HEWs received training in CBN	94.8	98.2
% of EAs where HEWs received *refresher* training	83.3	82.1
VCHWs trained in CBN		
% of EAs where VCHWs received *refresher* training	82.1	77.4
Additional HEW training		
% of EAs where HEWs received training in ENA [a]	54.7	42.9
% of EAs where HEWs received training in EOS [b]	69.0	70.2
% of EAs where HEWs received training in additional food support	62.1	59.6
% of EAs where HEWs received training in safety net	32.1	12.3

[a] Essential Nutrition Actions
[b] Enhanced Outreach Strategy

Table 3.3: Changes in child weighing (%), by location

	Tranche 2				Tranche 3			
	Baseline	Midline	N	Sig	Baseline	Midline	N	Sig
% Children weighed in the last 3 months	14.3	33.2	60	***	7.4	30.4	61	***
If yes, reported location of weighing:								
Health facility	78.2	49.2	41	***	88.7	45.1	30	***
Community weighing session	21.8	50.8	41	***	11.4	54.9	30	***
Out of total population								
% of total population weighed at health facility	11.4	15.4	60		6.2	10.7	61	**
% of total population weighed at community weighing session	2.6	16.4	60	***	1.1	19.3	61	***

* = p<=0.05; ** = p<=0.01; *** = p<0.001

Table 3.4: Distribution by age of children weighed total and by location (%)

	Tranche 2				Tranche 3			
	Baseline	Midline	N	Sig	Baseline	Midline	N	Sig
% Children weighed in the last 3 months								
0-11 months	12.8	22.9	59	*	8.7	29.5	59	***
12-23 months	16.6	40.6	59	***	8.0	36.0	60	***
24-35 months	13.0	37.4	58	***	5.4	26.5	59	***
% Children weighed at health facility, out of total population								
0-11 months	11.4	10.2	58		7.7	14.6	59	*
12-23 months	12.6	17.5	59		6.5	10.7	60	
24-35 months	10.5	17.8	59		4.1	8.0	59	
% Children weighed at community weighing session, out of total population								
0-11 months	1.5	12.7	59	***	0.2	14.5	59	***
12-23 months	3.6	21.0	59	***	1.5	25.0	60	***
24-35 months	2.3	17.4	58	***	1.3	18.4	59	***

* = p<=0.05; ** = p<=0.01; *** = p<0.001

Table 3.5: Change in possession and type of health card (%)

	Tranche 2				Tranche 3			
	Baseline	Midline	N	Sig	Baseline	Midline	N	Sig
Has Family or Other Health Card	50.5	64.0	60		45.8	60.7	61	***
If yes, has health card, which type:								
Family Health Card (FHC) only	21.6	40.9	59	***	30.1	35.0	58	
Other Health Card (OHC) only:	67.6	31.1	59	***	52.3	39.2	58	*
Has both FHC & OHC	10.9	28.0	59	***	17.6	25.9	58	
If yes has card, Health Card is kept:								
In household	96.8	91.5	59	**	91.1	93.3	58	
In clinic	3.2	8.5	59	**	8.9	6.7	58	
Out of total population:								
% of population with FHC only	11.2	23.4	60	***	13.7	21.6	61	*
% of population with OHC only	33.3	21.5	60	**	23.3	22.1	61	
%of population with both FHC & OHC	6.0	19.1	60	***	8.8	16.9	61	*

* = p<=0.05; ** = p<=0.01; *** = p<0.001

Table 3.6: Change in number and location of contacts with HEWs in the 6 months prior to the survey, and information received (%)

	Tranche 2				Tranche 3			
	Baseline	Midline	N	Sig	Baseline	Midline	N	Sig
Reported number of times had contact with HEW in last 6 months								
No contact	23.1	15.1	60	*	33.3	16.8	61	***
1-3 times	76.1	45.9	60	***	45.8	42.8	61	
4-6 times	0.8	27.9	60	***	18.0	34.1	61	***
More than 6 times	0.0	11.1	60	***	3.0	12.9	61	***
Reported location of contact with HEW								
Health post	40.9	76.2	60	***	37.1	68.6	61	***
Community outreach	53.4	63.5	60	*	43.8	54.2	61	**
House visit	45.0	66.8	60	***	34.8	58.3	61	***
Community conversation	29.5	59.4	60	***	22.0	52.0	61	***
Growth monitoring program	28.1	47.7	60	***	17.5	39.0	61	***
Model family training	18.0	30.4	60	**	12.4	28.2	61	***
Received the following information from HEWs								
Child weight/growth	45.9	65.9	60	***	34.6	57.3	61	***
Complementary feeding	50.9	79.7	60	***	43.0	68.6	61	***
Family planning	67.0	83.5	60	***	58.0	75.6	61	***
Child caring practices	56.7	76.0	60	***	46.3	67.6	61	***

* = p<=0.05; ** = p<=0.01; *** = p<0.001

Table 3.7: Change in number and location of contacts with VCHWs in the 6 months prior to the survey, and information received (%)

	Tranche 2				Tranche 3			
	Baseline	Midline	N	Sig	Baseline	Midline	N	Sig
Reported number of times had contact with VCHW in last 6 months								
No contact	46.0	28.8	60	***	64.8	36.1	61	***
1-3 times	34.4	34.7	60		23.6	30.9	61	
4-6 times	14.5	24.5	60	***	8.7	19.0	61	***
More than 6 times	5.2	12.0	60	**	2.9	14.0	61	***
Reported location of contact with VCHW								
Community outreach	36.1	52.2	60	***	20.2	39.4	61	***
House visit	38.0	61.8	60	***	23.0	51.9	61	***
Community conversation	25.8	54.7	60	***	17.0	43.5	61	***
Growth monitoring program	22.1	39.4	60	***	12.7	31.8	61	***
Received the following information from VCHWs								
Child weight/growth	31.5	54.0	60	***	14.6	41.7	61	***
Exclusive breastfeeding	33.0	56.7	60	***	21.0	41.7	61	***
Complementary feeding	33.9	57.0	60	***	19.8	45.7	61	***
Family planning	57.0	41.3	60	***	26.4	51.0	61	***
Child caring practices	35.6	55.9	60	***	23.0	47.3	61	***

* = p<=0.05; ** = p<=0.01; *** = p<0.001

Table 3.8: Summary of program coverage and participation estimates at the midline survey (%)

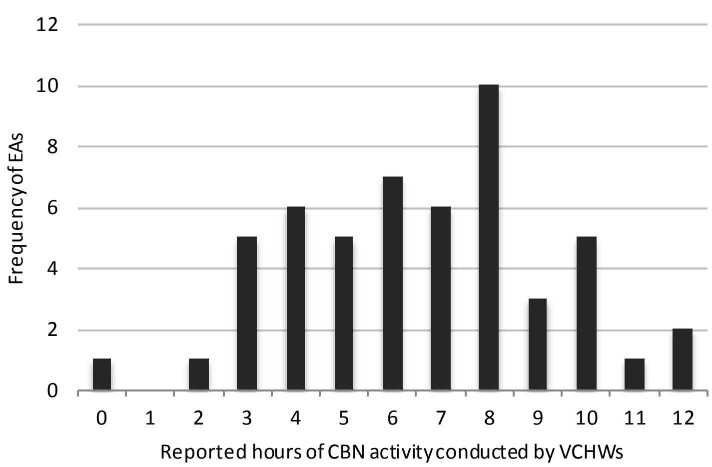
	Tranche 2	Tranche 3
Coverage/Participation Indicators		
Training (refresher) of VCHWs in CBN by midline (%EAs) (table 3.2)	82	77
Child participation in community weighing in last 3 months (table 3.3)	16	19
Child participation in community or health facility weighing in last 3 months (table 3.3)	33	30
Possession of Family Health Card (table 3.5)	64	61
Contact with HEW in last 6 months (table 3.6):		
Once or more	75	73
> Three times	39	47
Contact with VCHW in last 6 months (table 3.7):		
Once or more	71	64
> Three times	37	33

Table 3.9: Estimates of program intensity (%)

	Tranche 2	Tranche 3
Program intensity		
% of EAs with greater than 7 hours spent by VHCWs on CBN activities weekly (high intensity)	52	50 [a]
% of EAs with an average of less than 20 children (0-2 years) per VCHW	58	52
VCHW intensity (number of VCHWs per children under-2)	1:18	1:19 [b]
% of EAs with any reported CBN routine data at midline survey	100	72

[a] Selecting out for CBN reporting EAs only
[b] Selecting out for CBN reporting EAs only *and* eliminating two outliers of 1:177

Figure 3.2: Frequency of hours of VCHW activity per week in *tranche 2*



`

Table 3.10: Infant and young child feeding (IYCF) indicators: levels and changes between baseline and midline (%)

	Tranche 2				Tranche 3			
	Baseline	Midline	N	Sig	Baseline	Midline	N	Sig
IYCF Indicators								
Early initiation of breastfeeding	52.5	56.4	60		56.9	59.4	60	
Exclusive breastfeeding under-6 months	66.6	88.9	50	***	85.3	79.4	48	
Continued breastfeeding at 1 year	99.5	99.5	48		92.2	94.2	49	
Introduction to solid, semi-solid, or soft foods	67.4	48.0	33	*	48.3	59.2	39	
Minimum dietary diversity between 6-23 months	27.4	49.7	60	***	31.9	48.8	60	***
Minimum meal frequency in breastfed children 6-23 months	64.8	66.8	60		57.9	66.0	60	*
Minimum acceptable diet in breastfed children 6-23 months	21.2	43.4	59	***	27.6	37.4	60	**

* = p<=0.05; ** = p<=0.01; *** = p<0.001

Figure 3.3: Distribution of children aged 6-23 months who received food from seven different food groups the previous day, *tranche 2* (* = p<0.05; ** = p<0.01; *** = p<0.001)

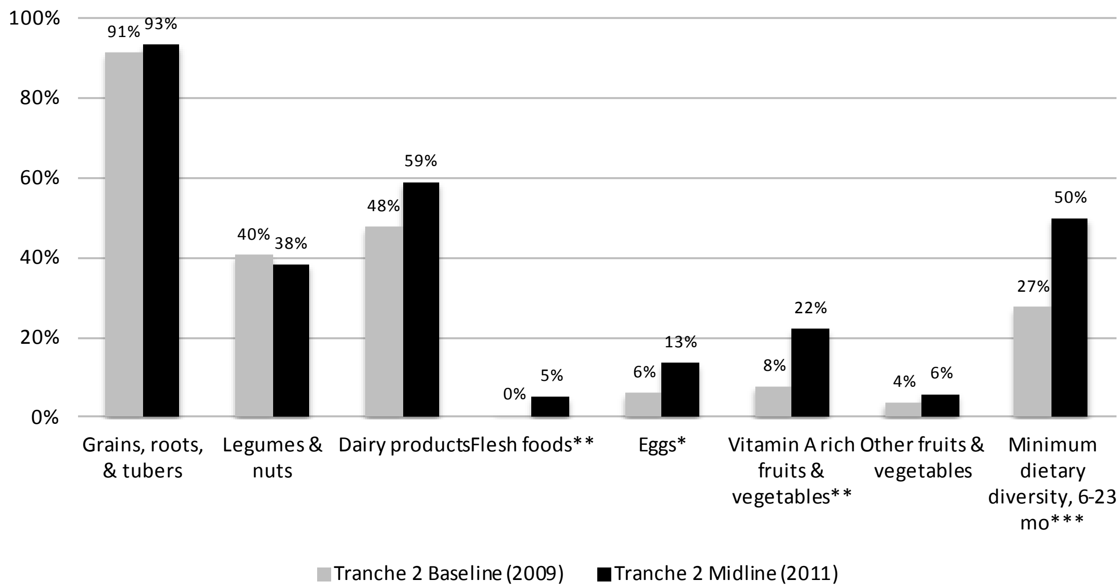


Figure 3.4: Distribution of children aged 6-23 months who received food from seven different food groups the previous day, *tranche 3* (* = p<0.05; ** = p<0.01; *** = p<0.001)

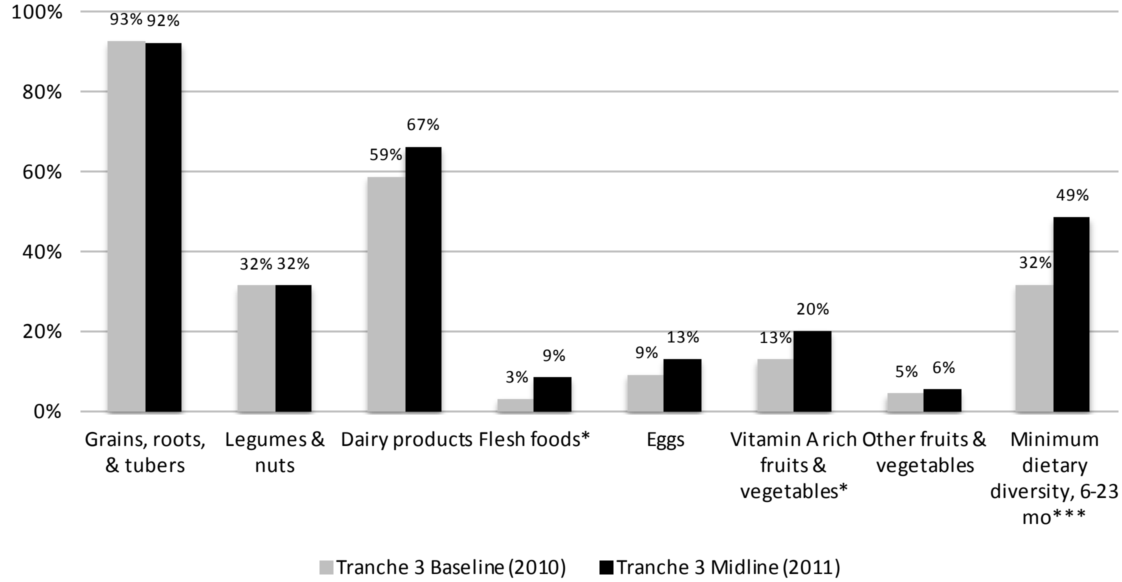


Table 3.11: Incidence and treatment of diarrhea (%)

	Tranche 2				Tranche 3			
	Baseline	Midline	N	Sig	Baseline	Midline	N	Sig
% Children with reported diarrhea in the last two weeks	30.3	20.7	60	***	27.0	21.7	61	*
Among children with diarrhea:								
Reported treatment								
% given ORS	23.7	37.7	56	*	24.1	34.7	54	*
% given government recommended homemade fluid "atimit"	18.3	21.6	56		25.0	32.1	54	
Reported fluid intake								
% given much less than usual or none	41.3	30.7	55		55.2	29.4	54	***
% about the same (or somewhat less)	32.6	50.6	55	**	27.2	50.1	54	**
% given more than usual	26.1	18.7	55		17.7	20.5	54	
Food intake								
% given none	15.0	22.3	55		16.5	23.7	54	
% given much less than usual	54.7	29.8	55	***	55.3	27.5	54	***
% given somewhat less usual	24.1	22.6	55		21.3	10.6	54	*
% given about the usual	4.4	19.2	55	**	6.2	32.9	54	***
% given more than usual	1.8	6.2	55		0.7	5.2	54	**

* = p<=0.05; ** = p<=0.01; *** = p<0.001

Table 3.12: Trends in stunting, underweight, and wasting (WHO Standards, <-2 SDs), 2000 to 2011 from DHS surveys; also comparing tranche 2 and 3 baseline prevalences

	DHS [a]			Tranche 2 Baseline	Tranche 3 Baseline	Change in ppts/yr in DHS surveys		
Prevalence	2000	2005	2010	2009	2010	2000-2005	2005-2011	2000-2011
Stunting	52.4	45.8	38.2	50.5	42.9	-1.3	-1.3	-1.3
Underweight	39.6	30.0	26.0	32.4	28.0	-1.9	-0.7	-1.2
Wasting	16.0	14.2	12.1	9.6	10.9	-0.4	-0.4	-0.4

[a] Selecting out for children under-3 and Amhara, Oromia, Tigray, and SNNPR only

Table 3.13a: Baseline and midline estimates of z-scores and prevalence for stunting, underweight, and wasting (WHO Standards, <-2 SDs), including significance level of change and overall change in term of percentage point change per year (%/yr) and 95% confidence intervals (CIs), *in tranche 2*

	Tranche 2						
	Baseline	Midline	N	Sig	Diff	CIs	ppts/yr
Stunting							
Height-for-age z-scores	-1.731	-1.473	60	*			
Stunting prevalence	50.5	40.6	60	***	-9.9	(-14.8 to -4.9)	-4.3/yr
Underweight							
Weight-for-height z-scores	-1.359	-1.285	60	0.335			
Underweight prevalence	32.4	28.7	60	0.119	-3.6	(-8.2 to +1.0)	-1.6/yr
Wasting							
Weight-for-height z-scores	-0.602	-0.645	60	0.554			
Wasting prevalence	9.3	15.4	60	**	+6.0	(+2.3 to +9.8)	+2.7/yr

* = p<=0.05; ** = p<=0.01; *** = p<0.001

Table 3.13b: Baseline and midline estimates of z-scores and prevalence for stunting, underweight, and wasting (WHO Standards, <-2 SDs), including significance level of change and overall change in term of percentage point change per year (%/yr) and 95% confidence intervals (CIs), *in tranche 3*

	Tranche 3						
	Baseline	Midline	N	Sig	Diff	CIs	ppts/yr
Stunting							
Height-for-age z-scores	-1.612	-1.470	60	0.107			
Stunting prevalence	42.9	38.5	60	0.114	-4.4	(-9.9 to +1.1)	-2.9/yr
Underweight							
Weight-for-height z-scores	-1.281	-1.282	60	0.989			
Underweight prevalence	28.0	28.7	60	0.788	+0.6	(-3.9 to +5.1)	+0.5/yr
Wasting							
Weight-for-height z-scores	-0.532	-0.599	60	0.326			
Wasting prevalence	10.8	10.6	60	0.919	-0.2	(-3.5 to +3.1)	-0.1/yr

* = p<=0.05; ** = p<=0.01; *** = p<0.001

Table 3.14a: Baseline and midline estimates of z-scores and prevalence for stunting, underweight, and wasting (WHO Standards, <-2 SDs), including significance level of change and overall change in term of percentage point change per year (%/yr) and 95% confidence intervals (CIs), *in reporting woredas in tranche 3*

	Tranche 3: CBN reporting EAs						
	Baseline	Midline	N	Sig	Diff	CIs	ppts/yr
Stunting							
Height-for-age z-scores	-1.734	-1.426	43	**			
Stunting prevalence	46.3	38.3	43	*	-8.0	(-14.7 to -1.3)	-5.3/yr
Underweight							
Weight-for-height z-scores	-1.377	-1.301	43	0.287			
Underweight prevalence	31.2	29.5	43	0.555	-1.7	(-7.5 to +4.1)	-1.3/yr
Wasting							
Weight-for-height z-scores	-0.580	-0.627	43	0.536			
Wasting prevalence	11.2	10.5	43	0.708	-0.7	(-4.7 to +3.2)	-0.5/yr

* = p<=0.05; ** = p<=0.01; *** = p<0.001

Table 3.14b: Baseline and midline estimates of z-scores and prevalence for stunting, underweight, and wasting (WHO Standards, <-2 SDs), including significance level of change and overall change in term of percentage point change per year (%/yr) and 95% confidence intervals (CIs), *in non-reporting woredas in tranche 3*

	Tranche 3: CBN non-reporting EAs						
	Baseline	Midline	N	Sig	Diff	CIs	ppts/yr
Stunting							
Height-for-age z-scores	-1.297	-1.568	16	*			
Stunting prevalence	34.1	38.5	16	0.339	+4.4	(-5.1 to +14.0)	+2.9/yr
Underweight							
Weight-for-height z-scores	-1.044	-1.237	16	*			
Underweight prevalence	20.1	27.3	16	*	+7.2	(+1.3 to +13.2)	+4.8/yr
Wasting							
Weight-for-height z-scores	-0.428	-0.546	16	0.451			
Wasting prevalence	9.8	11.2	16	0.685	+1.4	(-5.6 to +8.3)	+0.9/yr

* = p<=0.05; ** = p<=0.01; *** = p<0.001

Table 3.14c: Baseline and midline estimates of z-scores and prevalence for stunting, underweight, and wasting (WHO Standards, <-2 SDs), including significance level of change and overall change in term of percentage point change per year (%/yr) and 95% confidence intervals (CIs), *in EAs with high VCHW activity in tranche 2*

	Tranche 2: High VCHW activity						
	Baseline	Midline	N	Sig	Diff	CIs	ppts/yr
Stunting							
Height-for-age z-scores	-1.836	-1.422	27	*			
Stunting prevalence	54.5	38.8	27	**	-15.6	(-23.8 to -7.5)	-6.8/yr
Underweight							
Weight-for-height z-scores	-1.460	-1.259	27	0.102			
Underweight prevalence	35.1	29.4	27	0.082	-5.7	(-12.3 to +0.8)	-2.5/yr
Wasting							
Weight-for-height z-scores	-0.588	-0.668	27	0.430			
Wasting prevalence	7.8	13.6	27	*	+5.8	(+0.8 to +10.7)	+2.5/yr

* = p<=0.05; ** = p<=0.01; *** = p<0.001

Table 3.14d: Baseline and midline estimates of z-scores and prevalence for stunting, underweight, and wasting (WHO Standards, <-2 SDs), including significance level of change and overall change in term of percentage point change per year (%/yr) and 95% confidence intervals (CIs), *in EAs with low VCHW activity tranche 2*

	Tranche 2: Low VCHW activity						
	Baseline	Midline	N	Sig	Diff	CIs	ppts/yr
Stunting							
Height-for-age z-scores	-1.501	-1.463	25	0.784			
Stunting prevalence	44.0	41.2	25	0.429	-2.8	(-9.9 to +4.4)	-1.2/yr
Underweight							
Weight-for-height z-scores	-1.143	-1.284	25	0.238			
Underweight prevalence	27.2	29.4	25	0.529	+2.2	(-5.0 to +9.5)	-1.0/yr
Wasting							
Weight-for-height z-scores	-0.557	-0.630	25	0.591			
Wasting prevalence	8.6	18.1	25	**	+9.5	(+2.8 to +16.1)	+4.1/yr

* = p<=0.05; ** = p<=0.01; *** = p<0.001

Figure 3.5: Change in stunting and severe stunting prevalence (WHO Standards, <-2 SDs) between baseline and midline (solid line), compared to expected stunting trend based upon DHS historical data (dashed line, beginning at baseline estimate from evaluation survey), *in tranche 2*

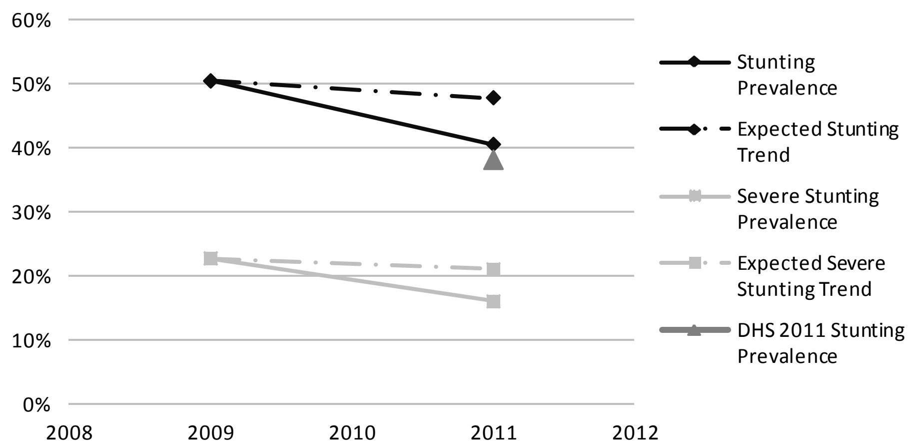


Figure 3.6: Change in stunting and severe stunting prevalence (WHO Standards, <-2 SDs) between baseline and midline (solid line), selecting out for CBN reporting EAs only, compared to expected stunting trend based upon DHS historical data (dashed line, beginning at baseline estimate from evaluation survey), *in tranche 3*

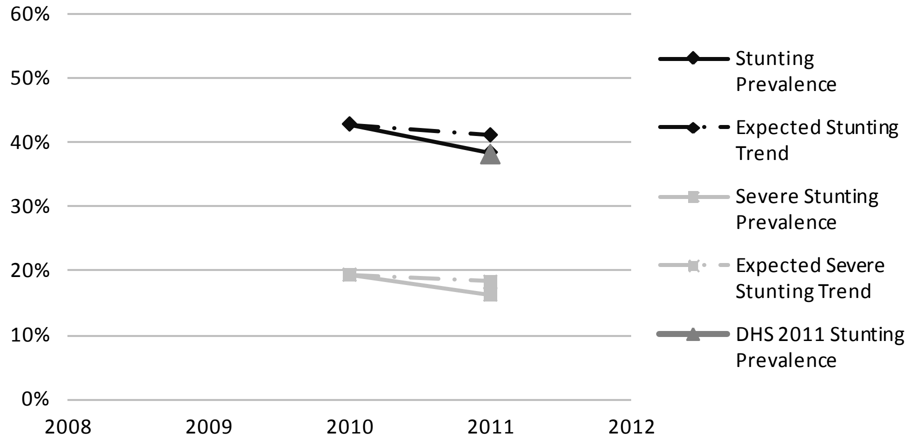


Figure 3.7: Long term DHS trend in stunting prevalence (WHO Standards, <-2 SDs) between 2000 and 2011 (selecting out for children under-3 and Amhara, Oromia, Tigray, and SNNPR only,) and trends in stunting prevalence from CBN evaluation surveys in tranche 2 and 3

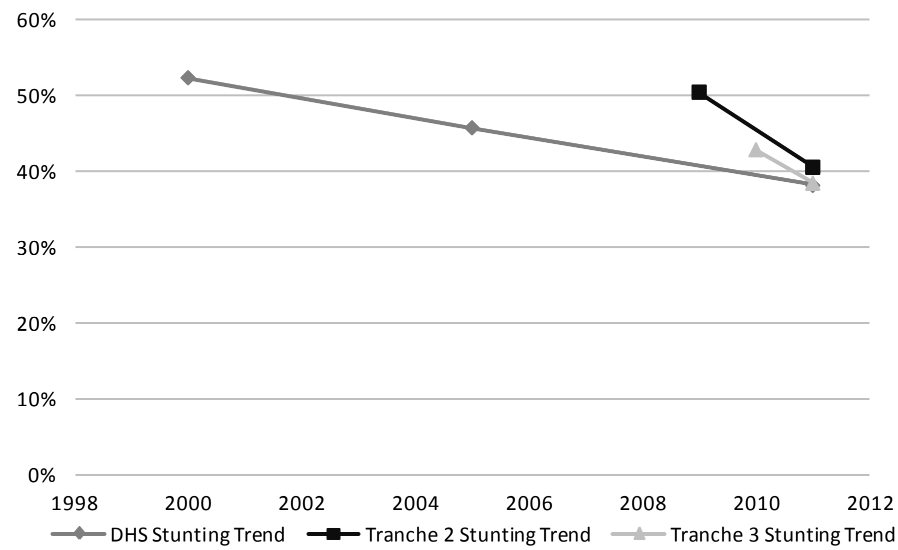


Table 3.15: Estimates of size of effect of CBN on HAZ and stunting prevalence

Tranche 2	Mean change in HAZ	Mean change in Stunting Prevalence (%)	N (EAs)	Tranche 3	Mean change in HAZ	Mean change in Stunting Prevalence (%)	N (EAs)
CBN implementation [a]				CBN implementation [b]			
Low CBN activity hours	+0.038	-2.8	25	No reporting	-0.270	+4.4	16
High CBN activity hours	+0.415	-15.6	27	Reporting	+0.307	-8.0	43
Effect size	+0.377	-12.8		Effect size	+0.577	-12.4	
Whole sample	+0.234	-9.5	52	Whole sample	+0.151	-4.6	59
p-value	0.115	0.020		p-value	0.003	0.046	
Effect size controlling	+0.377	-11.2		Effect size controlling	-0.580	-13.8	

[a] Calculated for tranche 2 using an indicator of the amount of time VCHWs spent per week on CBN activities (house-to-house visits, counseling, teaching, community discussion).
[b] Calculated for tranche 3, a comparison reporting versus non-reporting (from routine data) CBN EAs.

Table 3.16: Change in level of caregiver education between baseline and midline surveys (%)

	Tranche 2				Tranche 3			
	Baseline	Midline	N	Sig	Baseline	Midline	N	Sig
Improved education								
Informal	1.7	2.5	60		2.0	3.5	61	
Pre-school	0.3	0.7	60		0.1	1.9	61	**
Primary	18.7	20.9	60		19.3	22.3	61	
Secondary	3.7	5.0	60		4.4	7.6	61	*
Higher	0.3	0.2	60		0.4	0.2	61	
Poor education								
No school	75.4	70.6	60		73.9	64.5	61	***
% any education	24.6	29.4	60		26.1	35.5	61	***
% no education	75.4	70.6	60		73.9	64.5	61	***

* = p<=0.05; ** = p<=0.01; *** = p<0.001

Table 3.17: Change in quality of roofing material between baseline and midline surveys (%)

	Tranche 2				Tranche 3			
	Baseline	Midline	N	Sig	Baseline	Midline	N	Sig
Improved roofing								
Corrugated iron sheet	30.8	43.0	60	***	32.9	44.9	61	***
Poor roofing								
Thatch or grass	53.4	44.2	60	***	59.6	51.7	61	**
Wood and mud	6.5	8.7	60		1.9	2.6	61	
Mud and stone	5.6	3.7	60		2.1	0.1	61	
Reed and bamboo	0.2	0.1	60		0.6	0.3	61	
Brick tiles	-	-	-		-	-	-	
Other	1.2	0.3	60		2.9	0.5	61	
% improved roofing	30.8	43.0	60	***	33.5	45.1	61	***
% poor roofing	69.1	56.9	60	***	66.5	54.9	61	***

* = p<=0.05; ** = p<=0.01; *** = p<0.001

Table 3.18: Change in toilet facility between baseline and midline surveys (%)

	Tranche 2				Tranche 3			
	Baseline	Midline	N	Sig	Baseline	Midline	N	Sig
Improved toilet facility								
Flush/pour toilet	-	0.2	60		-	-	-	
Ventilated improved pit latrine	0.8	2.0	60		0.2	0.4	61	
Pit latrine with slab	1.2	18.7	60	***	0.8	9.8	61	***
Composting toilet	5.4	2.2	60	*	3.3	7.4	61	
Poor toilet facility								
Pit latrine without slab/open pit	57.6	57.5	60		60.4	52.9	61	
No facilities/bush	28.0	15.7	60	***	29.5	25.1	61	
Disposing on farm	5.4	0.6	60	**	5.6	2.2	61	**
Other	1.7	4.0	60		0.3	2.1	61	
% improved toilet facilities	7.3	23.0	60	***	4.3	18.2	61	***
% poor toilet facilities	92.7	77.0	60	***	95.7	81.8	61	***

* = p<=0.05; ** = p<=0.01; *** = p<0.001

Table 3.19: Change in drinking water source between baseline and midline surveys (%)

	Tranche 2				Tranche 3			
	Baseline	Midline	N	Sig	Baseline	Midline	N	Sig
Improved drinking water source								
Piped into dwelling	0.0	0.4	60		-	-	-	
Piped into yard or plot	0.5	0.0	60		-	-	-	
Public tap/standpipe	17.7	34.6	60	**	18.1	25.9	61	*
Tube well/borehole	18.2	13.9	60		14.1	11.8	61	
Protected well	7.0	1.4	60	**	6.6	2.0	61	**
Protected spring	16.7	15.8	60		12.5	13.6	61	
Rainwater collection	0.0	0.1	60		3.0	0.6	61	
Poor drinking water source								
Unprotected well	3.7	1.2	60	*	5.0	4.5	61	
Unprotected spring	29.2	22.0	60	*	25.1	29.4	61	
Tanker truck	-	-	-		0.0	0.2	61	
Surface water	7.1	10.2	60		13.7	11.8	61	
Other	0.0	0.5	60		0.6	0.3	61	
% improved drinking water source	50.0	66.3	60		56.2	54.1	61	
% poor drinking water source	40.0	33.7	60		43.8	45.9	61	

* = p<=0.05; ** = p<=0.01; *** = p<0.001

Table 3.20: Change in use and location of Antenatal Care between baseline and midline (%)

	Tranche 2				Tranche 3			
	Baseline	Midline	N	Sig	Baseline	Midline	N	Sig
ANC								
% women reported any ANC during last pregnancy	54.2	68.9	60	***	56.2	62.1	61	*
Frequency of ANC								
Once	10.2	7.4	59		10.5	6.8	61	
2-3 visits	57.8	57.2	59		51.2	53.0	61	
4 or more visits	32.0	35.5	59		38.3	40.3	61	
Location of ANC								
Hospital or clinic [a]	6.7	6.6	59		3.9	4.5	61	
Health Center	41.6	39.0	59		35.0	47.5	61	**
Health post	41.1	64.2	59	***	57.6	55.9	61	
Other	2.5	0.8	59		0.2	1.4	61	

[a] Government hospital; private hospital; government clinic; private clinic; or NGO clinic/health center

* = p<=0.05; ** = p<=0.01; *** = p<0.001

Table 3.21: Change in reported Antenatal Care services received between baseline and midline (%)

	Tranche 2				Tranche 3			
	Baseline	Midline	N	Sig	Baseline	Midline	N	Sig
ANC services reported received								
Weighed	52.9	63.5	59	*	57.3	72.7	61	***
Blood pressure measured	58.6	64.0	59		53.6	69.1	61	**
Urine tested	15.8	32.0	59	***	16.6	30.2	61	***
Blood tested	25.0	46.0	59	***	23.3	37.9	61	***
Malaria drug given	15.4	30.9	59	***	18.2	20.4	61	
Counseling reported received during ANC on:								
Breastfeeding	49.0	76.4	59	***	46.7	60.0	61	**
Family planning	73.2	85.3	59	**	67.6	75.2	61	
HIV/AIDS	64.7	81.5	59	***	64.0	74.5	61	**
Maternal nutrition	60.5	79.3	59	***	59.1	68.0	61	

* = p<=0.05; ** = p<=0.01; *** = p<0.001

Table 3.22: Change in health care and nutrition during pregnancy between baseline and midline (%)

	Tranche 2				Tranche 3			
	Baseline	Midline	N	Sig	Baseline	Midline	N	Sig
Iron intake								
% women reported taking iron/folate during last pregnancy	28.5	49.6	59	****	30.7	39.4	61	*
If taken, number of days consumed								
1-7	100.0	25.3	40	***	52.8	27.8	43	**
8-15	0.0	17.7	40	***	22.6	19.9	43	
16-30	0.0	39.2	40	***	9.5	42.3	43	***
31+	0.0	17.9	40	***	2.9	9.9	43	*
Tetanus injection								
% women reported receiving tetanus injection during last pregnancy	82.3	64.1	59	***	77.1	62.3	60	***
Number of tetanus injections received								
1	82.3	23.6	59	***	21.8	26.0	59	
2-3	0.0	69.6	59	***	64.8	68.5	59	
4-5	0.0	6.8	59	***	13.4	5.6	59	**
De-worming								
% women reported receiving de-worming tablet during last pregnancy	12.3	12.7	60		11.7	10.5	61	**

* = p<=0.05; ** = p<=0.01; *** = p<0.001

Table 3.23: Change in reported food intake during pregnancy between baseline and midline (%)

	Tranche 2				Tranche 3			
	Baseline	Midline	N	Sig	Baseline	Midline	N	Sig
Reported food intake during pregnancy								
Less than usual	64.3	46.9	60	***	63.4	55.0	61	**
Same as usual	27.3	45.1	60	***	30.6	39.5	61	**
More than usual	8.4	8.1	60		6.0	5.5	61	

* = p<=0.05; ** = p<=0.01; *** = p<0.001

Table 3.24: Change in delivery care between baseline and midline (%)

	Tranche 2				Tranche 3			
	Baseline	Midline	N	Sig	Baseline	Midline	N	Sig
Delivery location								
Home delivery [a]	96.5	92.1	60	**	94.8	92.7	61	
Institutional delivery [b]	3.5	6.9	60	*	5.2	7.3	61	
Delivery assistance								
Health Professional [c]	3.4	6.6	60	*	5.2	6.9	61	
Health Extension Worker	0.9	5.0	60	***	1.1	2.8	61	*
Traditional Birth Attendant	38.2	33.7	60		37.8	30.1	61	*
Community Health Worker	1.0	3.0	60	**	1.8	1.6	61	
Relative/Friend	49.2	59.5	60	**	46.0	60.5	61	**
None	3.4	2.0	60		4.8	1.1	61	
Other	6.5	5.6	60		6.7	3.2	61	
Delivery care for child								
Child weighed at birth	8.3	14.8	60	**	7.3	17.8	61	***

[a] Own home; mother's home
[b] Government hospital; government clinic/health center; private hospital; private clinic; maternal home; NGO clinic; other public
[c] Doctor; nurse; midwife; auxiliary midwife

* = p<=0.05; ** = p<=0.01; *** = p<0.001

Table 3.25: Change in post-delivery care between baseline and midline (%)

	Tranche 2				Tranche 3			
	Baseline	Midline	N	Sig	Baseline	Midline	N	Sig
Post-delivery visit								
% women visit by health professional after delivery	21.5	29.8	60	**	17.8	29.0	61	
Post-delivery assistance								
Health Extension Worker	69.0	70.6	49		72.8	68.5	46	
Community Health Worker	23.7	35.8	49	*	26.0	23.4	46	
Traditional Birth Attendant	0.6	5.5	49	*	6.9	5.0	46	
Other health worker	5.5	2.7	49		1.4	9.0	46	*
Timing of post-delivery visit								
Within 1 hour of birth	8.7	27.9	49	***	15.2	20.6	46	
Within 2-24 hours of birth	6.3	22.5	49	***	19.5	21.9	46	
Within 25-48 hours of birth	12.8	16.0	49		14.8	14.8	46	
More than 48 hours after birth	72.3	33.6	49	***	50.6	42.7	46	
Vitamin A post-delivery								
% women who received Vitamin A within 2 months of birth	24.1	40.2	60	***	18.7	30.0	61	***

* = p<=0.05; ** = p<=0.01; *** = p<0.001

Table 3.26: Children participating in nutritional screening for TSF (%)

	Tranche 2				Tranche 3			
	Baseline	Midline	N	Sig	Baseline	Midline	N	Sig
Child participated in recent nutritional screening	54.1	52.7	60		31.4	34.4	61	
If not, reason reported:								
Did not know about the date/time	17.9	48.4	55	***	26.5	37.6	58	*
Caretaker unable to take child	11.6	3.2	55	**	7.2	3.4	58	
Child sick	1.0	0.0	55	*	1.0	0.7	58	
Migrated	2.5	0.5	55		0.1	1.6	58	*
Distance to screening center too far	0.4	2.5	55		2.5	0.0	58	*
Busy with agricultural activities	0.7	3.1	55	*	1.1	3.0	58	
Other reason	48.1	18.6	55	***	40.8	25.0	58	*
Don't know why	17.8	23.6	55		20.8	28.8	58	
Child received supplementary food after the last screening	14.6	16.7	57		20.8	12.7	47	

* = p<=0.05; ** = p<=0.01; *** = p<0.001

Table 3.27: Mothers participating in nutritional screening for TSF (%)

	Tranche 2				Tranche 3			
	Baseline	Midline	N	Sig	Baseline	Midline	N	Sig
Mother reported:								
Participated in screening when pregnant with child	24.9	26.5	60		11.4	14.4	61	
Received supplementary food while pregnant with child	48.4	47.5	42		58.1	36.9	26	*
Received rations while child was <6 months of age and mother was breastfeeding	15.1	11.9	60		19.2	5.2	61	***
Received supplementary food while child was above 6 months	15.4	10.0	60		15.8	4.1	61	***

* = p<=0.05; ** = p<=0.01; *** = p<0.001

Table 3.28: Difference in HAZ and stunting prevalence estimates between baseline and midline, in relation to intensity of CBN implementation and receipt of supplementary food from TSF in EAs *in tranche 2*

	Height-for-Age z-score		Stunting Prevalence	
	No food received	Any food received	No food received	Any food received
Low activity intensity	-0.230 (11)	+0.248 (14)	+1.4 (11)	-6.1 (14)
High activity intensity	+0.154 (11)	+0.595 (16)	-16.0 (11)	-15.4 (16)

Figure 3.8: Change in stunting prevalence, controlling for intensity of CBN implementation and receipt of supplementary food from TSF in EAs *in tranche 2*

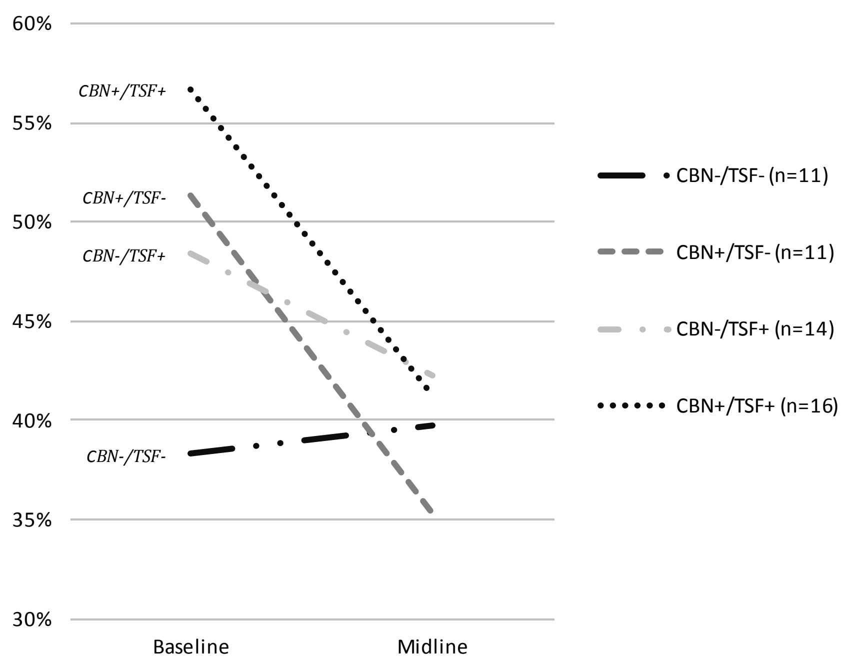


Table 3.29: Regression (OLS) coefficients (with p-value), dependent variables mean HAZ and mean stunting prevalences differences (midline-baseline), of CBN and receipt of supplementary food (TSF) controlling for changes in socioeconomic and food security indicators

	Mean HAZ difference		Mean stunting prevalence difference	
Independent variables	Tranche 2	Tranche 3	Tranche 2	Tranche 3
CBN activity	0.377	0.580	-0.112	-0.138
intensity/Reporting [a]	(0.151)	(0.005)	(0.070)	(0.036)
Receipt of supplementary	0.446	-0.251	-0.024	0.070
food	(0.076)	(0.183)	(0.678)	(0.255)
Education improved	0.259	-0.143	-0.031	-0.023
	(0.335)	(0.468)	(0.624)	(0.725)
Roofing improved	-0.127	0.006	0.017	-0.041
	(0.634)	(0.978)	(0.780)	(0.556)
Toilet improved	0.186	-0.013	-0.013	0.022
	(0.501)	(0.946)	(0.843)	(0.726)
Drinking water improved	0.124	0.215	-0.030	-0.092
	(0.635)	(0.266)	(0.618)	(0.144)
ANC improved	-0.026	0.083	-0.075	-0.042
	(0.930)	(0.643)	(0.273)	(0.473)
Little to no hunger	0.368	-0.648	-0.157	0.244
difference	(0.601)	(0.206)	(0.342)	(0.179)
Constant	-0.483	-0.253	0.077	0.134
Adj R Sq	0.005	0.102	0.005	0.047
N	52	59	52	59

[a] In Tranche 3, it is not CBN activity intensity, but rather CBN reporting included in regression. Since CBN in Tranche 3 had only been implemented for 6 months or less in the majority of surveyed CBN EAs, there was little difference in intensity of implementation between EAs.

Table 3.30: Household Hunger Scale classifications between baseline and midline (%)

	Tranche 2				Tranche 3			
	Baseline	Midline	N	Sig	Baseline	Midline	N	Sig
Little to no household hunger	80.9	88.1	60	**	86.3	82.7	61	
Moderate hunger in the household	15.7	10.8	60	*	12.4	13.7	61	
Severe hunger in the household	3.4	1.1	60	*	1.4	3.6	61	

* = p<=0.05; ** = p<=0.01; *** = p<0.001

Table 3.31: Proportion of Hot Spot Priority One woredas at baseline and midline surveys in each tranche compared to national average

	Around Baseline		Around Midline
Tranche 2	April 2009		June 2011
	57% (34/60)		33% (20/60)
Tranche 3	March 2010		June 2011
	31% (19/61)		31% (19/61)
National	April 2009	March 2010	June 2011
	28% (182/650)	32% (205/650)	22% (179/800)

Table 3.32: Seasonality of baseline and midline surveys in both tranches, by region

	Tranche 2		Tranche 3	
Region	Baseline Jun-Jul 2009	Midline Sept 2011	Baseline Mar-Apr 2010	Midline Sept 2011
Amhara	Hunger	End of hunger/harvest	Intermediate (short rains)	End of hunger/harvest
Oromia	Hunger	End of hunger/harvest	Intermediate (short rains)	End of hunger/harvest
SNNPR	End of hunger	Harvest	Hunger	Harvest
Tigray	Hunger	End of hunger/harvest	Intermediate (short rains)	End of hunger/harvest
National	2009 had poor early (belg) rains, and high food prices, so Jun-Jul may have higher than usual malnutrition in some areas	Drought started late 2010, belg season poor, severe problems in Somalia probably affected E Ethiopia; probably higher than usual malnutrition	2010 had good belg rains, food prices fell, nutrition probably improved	Same as tranche 2 midline
Likely effect	Seasonal change Jul-Sept probably less important than drought in 2011; slight worsening expected		Probably a relatively good nutrition situation at baseline; definite worsening expected from drought, base-midline	

Table 3.33: Summary of likely food security changes base - to mid-line

	Hunger scale	Hotspots	Seasonal	National drought
Tranche 2	Improving	Improving	Same	Worse slightly
Tranche 3	Worsens	Same	Possibly worse	Worse
See	*Table 3.30*	*Table 3.32*	*Table 3.33*	*Text*

Table 4.1: Process indicators: objectives and achievements

Objective/Indicator	Aim	Source	Found: Tranche 2 *(Baseline & Year 2)*	Found: Tranche 3 *(Baseline & Year 1)*	See table/figure	Score [a]
% Children 0-3 yrs weighed in given month [b]	Baseline: - Year 3: 50%	PIM, p 112	Baseline: 14% Midline: 33%	Baseline: 7% Midline: 30%	Table 3.3	+
% HEWs trained on nutrition (ENA)	Baseline: - Year 3: 60%	PIM, p 113	Baseline: - Midline: 55%	Baseline: - Midline: 43%	Table 3.2	+
% Pregnant women receiving iron/folate	Baseline: 10% Year 3: 30%	PIM, p 113	Baseline: 29% Midline: 50%	Baseline: 31% Midline: 39%	Table 3.22	++
% HEW and VCHW refresher training	Baseline: - Year 3: 40%	PIM, p 114	Baseline: - Midline: ~80%	Baseline: - Midline: ~80%	Table 3.2	++
VCHWs ratio to children 0-2 yrs	Target: 1:20 [c]	PIM, p 26	Baseline: - Midline: 1:18	Baseline: - Midline: 1:19	Text: '1 per 50 hhds'	++

[a] ++ Exceeding aim; + Improving or in line with aim; - Little to no improvement, below level aimed at
[b] The original target in PIM is for children under-2, but results data from CBN evaluation data is for children under-3
[c] Calculated by taking 40% (proportion of under-2 population) of the proposed VCHW ratio of 1 to every 50 households.

Table 4.2: Indicators of infant and young child feeding: objectives and achievements

Objective/Indicator	Aim	Source	Found: Tranche 2 *(Baseline & Year 2)*	Found: Tranche 3 *(Baseline & Year 1)*	See table/figure	Score [a]
Proportion of infants 0-6 months exclusively breast fed	Baseline: 32% Year 3: 54%	PIM, p 110	Baseline: 67% Midline: 89%	Baseline: 85% Midline: 79%	Table 3.10	++
Proportion of children who started breastfeeding within 1 hour of birth	Baseline: 69% Year 3: 83%	PIM, p 110	Baseline: 53% Midline: 56%	Baseline: 57% Midline: 59%	Table 3.10	-
Proportion of infants 6-9 months introduced to complementary food at 6-7 months	Baseline: 25% Year 3: 40%	PIM, p 110	Baseline: 67% Midline: 48%	Baseline: 48% Midline: 59%	Table 3.10	+
Proportion of infants 6-23 months with a minimum acceptable diet	Not included in PIM		Baseline: 21% Midline: 43%	Baseline: 28% Midline: 37%	Table 3.10	++
Proportion of children with diarrhea who were fed "same as or more than usual"	Baseline: 15% Year 3: 37%	PIM, p 110	Baseline: 6% Midline: 25%	Baseline: 7% Midline: 38%	Table 3.11	+

[a] ++ Exceeding aim; + Improving or in line with aim; - Little to no improvement, below level aimed at

Table 4.3: Differences in HAZ and stunting prevalence changes between internal comparison groups

	Comparison groups					
	VCHW activity high			VCHW activity low		
	Baseline	Midline	Difference (a)Total ppts (CIs) (b) Est. ppts/yr	Baseline	Midline	Difference (a)Total ppts (CIs) (b) Est. ppts/yr
Tranche 2						
Mean Height-for-age z-scores	-1.84	-1.42*	NA	-1.50	-1.46 [0]	NA
Stunting Prevalence (%)	54.5	38.8**	-15.7 (-23.8 to -7.5) -6.8 ppts/yr	44.0	41.2 [0]	-2.8 (-10.0 to +4.4) -1.2 ppts/yr
N: clusters (children)	27 (460)	27 (460)		25 (425)	25 (425)	
	CBN reporting			Non-CBN reporting		
	Baseline	Midline	Difference (a)Total ppts (CIs) (b) Est. ppts/yr	Midline	Baseline	Difference (a)Total ppts (CIs) (b) Est. ppts/yr
Tranche 3						
Mean Height-for-age z-scores	-1.73	-1.43**	NA	-1.30	-1.57*	NA
Stunting Prevalence (%)	46.3	38.3*	-8.0 (-14.7 to -1.3) -5.3 ppts/yr	34.1	38.5 [0]	+4.4 (-5.1 to +1.4) +4.9 ppts/yr
N: clusters (children)	43 (730)	43 (730)		16 (270)	16 (270)	

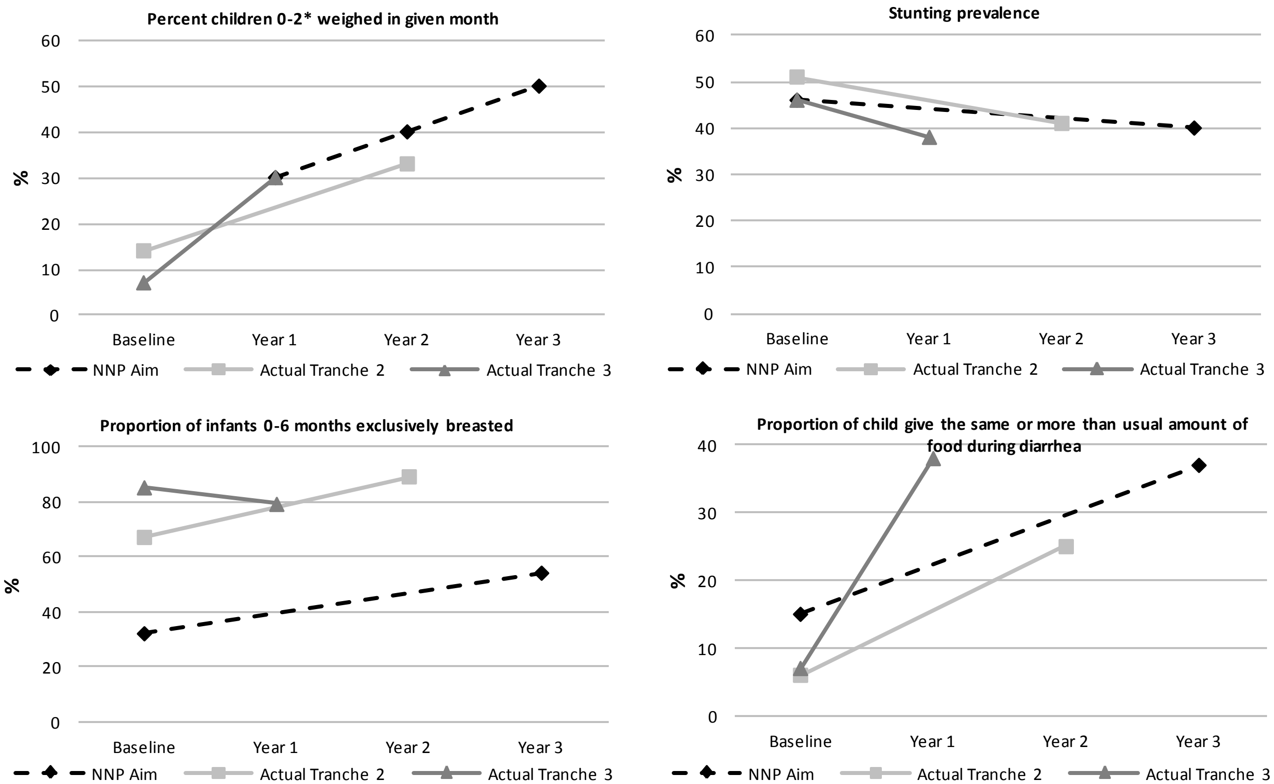
Differences base-to midline: ** p<0.01; * p<0.05; [0] Not significant
Notes: The same clusters as baseline were re-sampled at midline. The child numbers (last rows) are estimated from average numbers per cluster. CIs: 95% confidence intervals.

Table 4.4: Anthropometric outcome indicators: objectives and achievements

Indicator	Aim	Source	Found: Tranche 2 (Baseline & Year 2)	Found: Tranche 3 [b] (Baseline & Year 1)	See table	Score [a]
Stunting	Baseline: 46% Year 2: 44% Difference: -2.0 ppts	PIM, p 90	Baseline: 51% Year 2: 41% Difference: -9.9 ppts	Baseline: 46% Year 2: 38% Difference: -8.0 ppts	3.13a/b	++
Underweight	Baseline: 38% Year 2: 35% Difference: -3.4 ppts	PIM, p 90	Baseline: 32% Year 2: 29% Difference: -3.7 ppts	Baseline: 31% Year 2: 30% Difference: -1.7 ppts	3.13a/b	+
Wasting	Baseline: 11% Year 2: 9% Difference: -2.2 ppts	PIM, p 90	Baseline: 9% Year 2: 15% Difference: +6.1 ppts	Baseline: 11% Year 2: 11% Difference: -0.7 ppts	3.13a/b	-

[a] ++ Exceeding aim; + Improving or in line with aim; - Little to no improvement, below level aimed at
[b] Selecting out for reporting EAs only

Figure 4.1: Examples of performance indicators met or exceeded, compared to targets in the project plan (PIM)



ANNEX

Age Heaping and re-calculation

During exploratory analysis of the four CBN evaluation surveys, two baselines and two midlines, extreme age heaping was discovered when investigating child age distribution. Bar graphs of the age of child variable in each survey show heaping particularly at 12 and 24 months, as seen in figures 1a and 1b. Age heaping is common occurrence in surveys since caregivers often report approximate estimates of the child's age in years as opposed to exact months. Child age is an important factor in the calculation of anthropometric outcomes as mis-reporting of age may result in incorrect z-score and under or over estimation of the prevalence of stunting and underweight.

Figure 1a: Distribution of the age of child values in Tranche 2 evaluation surveys

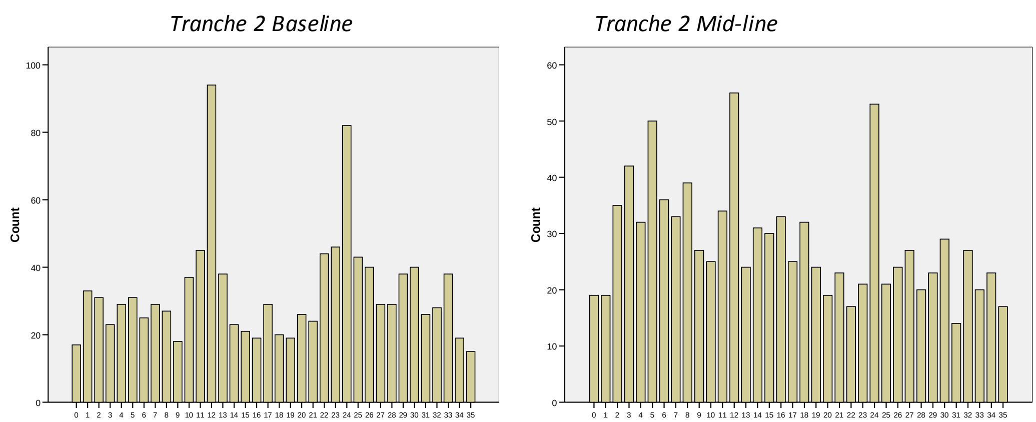
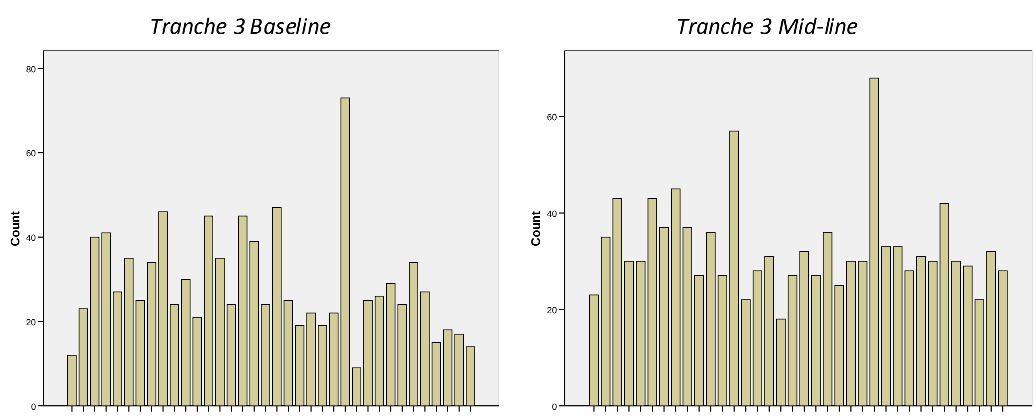


Figure 1b: Distribution of the age of child values in Tranche 3 evaluation surveys



The age of each child was queried using two separate questions during data collection, detailed in figure 2. The first estimation, question IC7, is intended to calculate age by subtracting the child's date of birth from the date of interview. This calculation is used in most large-scale surveys, including DHS, to estimate a child's age in months. The second estimation, question IC8, determines age by asking the child's the age of the child in months at their last birthday. This IC8 estimation of age allows for the possibility of rounding the child's age in months up or down to the nearest even year if the caregiver does not immediately know child's age in exact months as opposed to years.

Figure 2: Age questions within CBN questionnaire

IC7	IN WHAT MONTH AND YEAR WAS [NAME] BORN? TRY TO LOOK AT IMMUNIZATION CARDS/OTHER DOUMENTS. USE LOCAL CALENDARS IF DOCUMENTS ARE NOT AVAILABLE	Day	/____/_____/
		DK day	98
		Month	/____/_____/
		DK month	98
		Year	/____/____/___/___/
		DK year	98
IC8	HOW OLD WAS [NAME] AT HIS/HER LAST BIRTHDAY? RECORD AGE IN COMPLETED MONTHS	Age in completed months	/____/____/

During initial analysis following data entry, however, age of the child was calculated using results from question IC8, discussed above. Data from question IC7 on date, month, and year of birth was entered into each database, but an age calculation based upon IC7 results was not created. As seen above in figures 1a and 1b, bar graphs of the originally calculated IC8 derived age variables showed extreme age heaping in all surveys.

Analysis of trends across survey years using cross-sectional data requires the establishment of similarity across sample populations. To measure the extent of age heaping across all surveys, a metric was calculated to determine if any particular survey would not be comparable to the other. An "ideal" age distribution metric was created using data from DHS 2000 and 2005 surveys, where age heaping is minimal, for comparison with the metrics created from evaluation surveys. To create the metric, data was aggregated by child age in months, and the number of cases in each month of age was then regressed on age, with the unstandardized residuals saved in the aggregated file. The absolute value of the residuals was summed and then divided by the total number of cases, resulting in a measure of age heaping. The closer the metric is to 0, the less variation that exists between each reported age month.

Analysis on two DHS surveys showed low and comparable metrics of 0.114 in 2000 and 0.140 in 2005. Metrics were calculated on the age variables created using IC8, shown in table 1. The metrics reach as high as 0.324 in the tranche 2 baseline and 0.305 in the tranche 3 midline. The importance of this measure lies not only in the value itself, but its comparative value across survey years. Baseline and midline metrics for both tranche 2 and 3 are quite different, as seen in table 1, therefore age cannot be reasonably compared between baseline and midline surveys within each tranche.

Table 1: Calculated metrics on age in months for DHS 2000 and 2005, and baseline and midline surveys in Tranche 2 and 3

	DHS		**Tranche 2**		**Tranche 3**	
	2000	*2005*	*Baseline**	*Midline**	*Baseline**	*Midline**
Metric	0.114	0.140	0.324	0.220	0.192	0.305

*calculated by age variable derived from question IC8

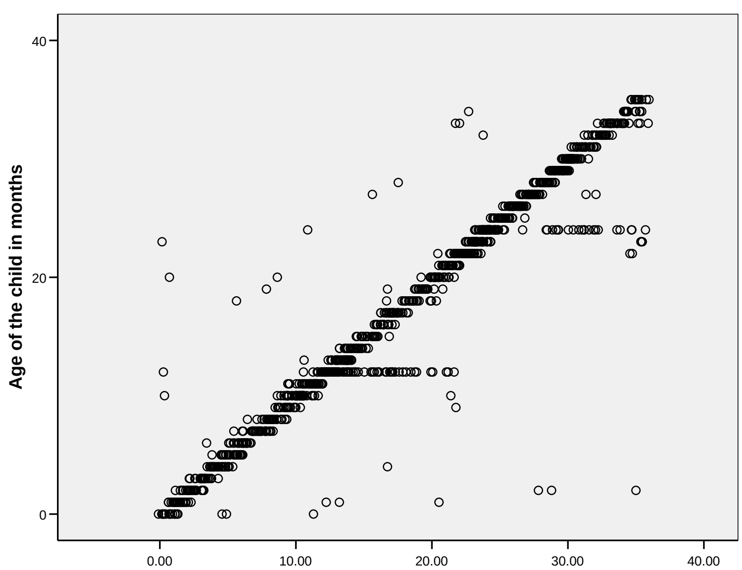
To remedy the heaping, age was re-calculated using data from responses to question IC7 on exact month and year of birth. Initial reports from data collectors on child age states that IC7 data was not used due to a large amount of missing data on “date of birth,” while question IC8 had 100% response rate. If age is calculated using only cases where a date of birth is available, around 10 to 20% of the sample is lost due to missing data (either responses of “I don’t know,” responses outside of the range of possible days in a month, or system missing).

Instead of losing this data due to unknown date of births, these missing cases can be reset to 15 (the middle of the month). In making this adjustment, the loss of sample size discussed above is avoided. By adjusting missing date of births to 15, there is only a risk of under-estimating the child’s age by 1 month. For example, if a child’s missing date of birth is changed to the 15th but they were actually born prior to the 15th in that month, their age will still be estimated accurately. If, though, a child’s date of birth is changed to the 15th but they were actually born after the 15th of that month, the child’s age may be underestimated by 1 month.

After cleaning the “outside of range” values and correcting for data-entry errors, adjusting the unknown date of birth responses to 15 resulted in higher response rates, between 97 and 98%, using age derived from question IC7 that are then comparable to the response rates of age derived from question IC8.

Scatterplots of the newly created IC7 age variable and the original IC8 age variable in each survey show evidence of possible age rounding down in IC8 responses, resulting in age heaping at 12 and 24 months. This is particularly evident in the tranche 2 baseline (see figure 3). When age is calculated as anywhere between 12 and 23 months in the new IC7 age variable (on the x-axis), it is often labeled simply as 12 in the IC8 variable (y-axis). The same is seen at 24 months with several cases calculated between 24 and 35 months using IC7 variables labeled as 24 in IC8. This is further evidence that the newly created IC7 age variables are more reliable than the IC8 variable.

Figure 3: Scatterplot of IC7 age (x-axis) and IC8 (y-axis)



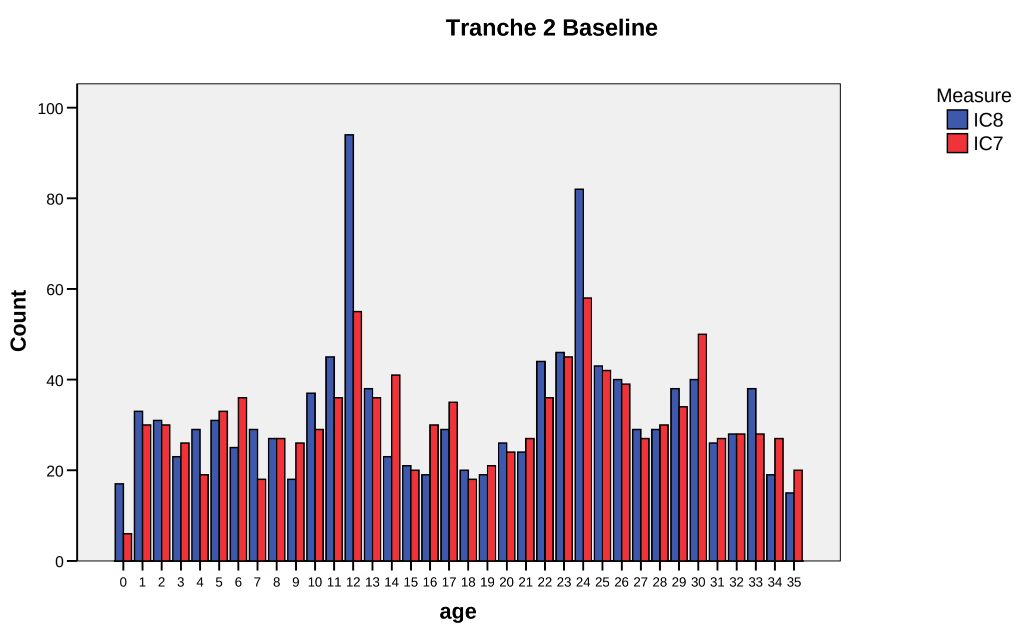
To further clean the age variable, data points in IC7 outside of a +2 or -2 range estimation in IC8 were eliminated since it could not be determined where the error was that caused the two age estimations to be so different (with the exception of cases where IC8 was equal to 12 or 24, since those errors can be attributed to age rounding down, as seen in figure 3). For example, there are several data points in the scatterplot above that appear to be around 12 months off of the common line between IC7 and IC8 age variables, most likely as a result an error in entry of the year of birth. Since, though, it cannot be determined for certain if that is the error present, these cases are eliminated.

New bar graphs of both IC7 and IC8 derived age show that the re-calculation of age using data from question IC7 greatly reduced age heaping at 12 and 24 months in all four surveys (see figure 4). Also, the metrics created using age derived from IC7 are not only lower (have less variation between months), but are more comparable between baseline and midline surveys in both tranches compared to those created from age derived from IC8, as seen in table 2.

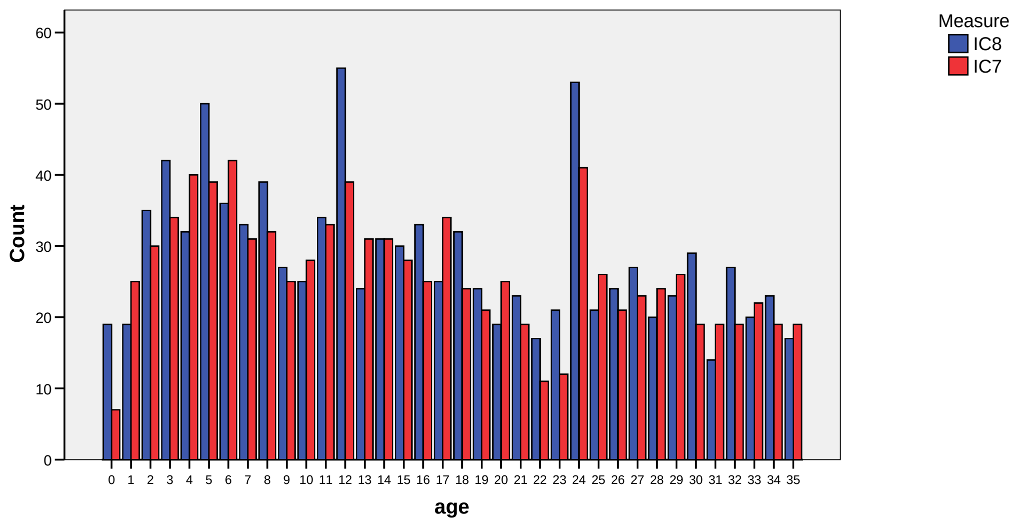
Table 2: Calculated metrics on age in months for DHS 2000 and 2005, and baseline and midline surveys in Tranche 2 and 3 using age derived from question IC7

	DHS		Tranche 2		Tranche 3	
	2000	*2010*	*Baseline*	*Midline*	*Baseline*	*Midline*
Metric (using IC7age)	0.114	0.140	0.261	0.194	0.204	0.244
Metric (using IC8 age)	-	-	0.324	0.220	0.192	0.305

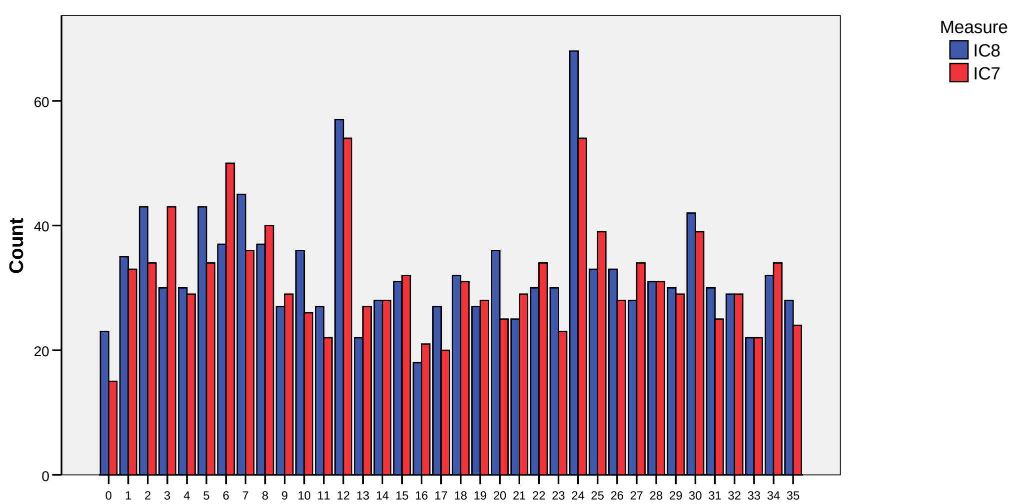
Figure 4: Comparison of IC8 and IC7 derived age in month variables



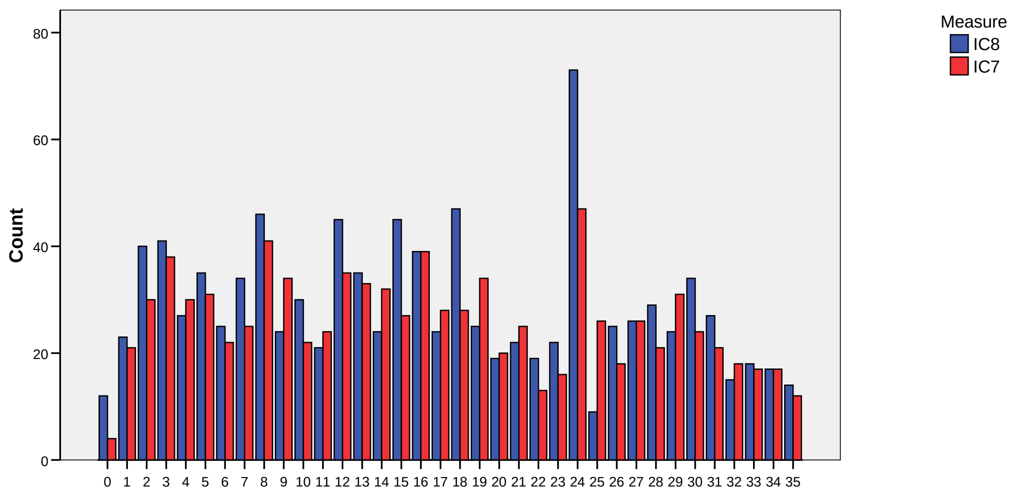
Tranche 2 Midline



Tranche 3 Midline



Tranche 3 Midline



The change in estimation of age also changed the anthropometric estimates when the newly calculated age variables using IC7 data were used to calculate new z-scores. Table 3 shows the change in stunting and underweight prevalence using the two different estimations of age. These estimates are based on WHO specified ranges (stunting: -6 to 6; underweight: -6 to 5). The reported prevalences used in the rest of this report are based upon z-scores with modified ranges to account for outliers. Wasting is not shown because age is not included in its calculation, therefore no change exists.

Table 3: Difference in stunting and underweight prevalence using two different age estimations

	Tranche 2		Tranche 3	
	Baseline	*Midline*	*Baseline*	*Midline*
Stunting Prevalence				
Previous estimation (based on IC8)	44.5%	37.5%	41.0%	36.4%
New estimation (based on IC7)	51.0%	41.3%	43.7%	39.2%
Underweight Prevalence				
Previous estimation (based on IC8)	28.4%	27.0%	26.2%	26.6%
New estimation (based on IC7)	33.4%	29.2%	28.5%	28.6%

Using the new estimate of age based on the IC7 data, prevalence of both stunting and underweight increases for all surveys, again evidence of age rounding down, which resulted in initial underestimation of the malnutrition burden.

Once cleaned, new z-score variables were created using the cleaned child age estimation. Further exploration of newly created z-scores resulted in a restricted range for height-for-age z-scores in all four surveys, reducing the range to values between -5 and +5, since few values existed outside the range and did not fall into the same distribution as the rest of the z-score values.

Following creation of a corrected age variable, the range of each anthropometric measure was explored using histograms and observation to determine appropriate ranges for each measure. As a result of identified outliers, height-for-age z-scores values outside of +5 and -5 were eliminated since few values existed outside the range and did not fall into the same distribution as the rest of the z-score values. Weight-for-age values were also restricted to a range of values above -5 for similar reasons. Weight-for-height values were left alone since no outliers were identified.

Bar graphs of the newly created age variables with adjusted ranges still showed some heaping at 12 and 24 months, and metrics slightly higher than the ideal DHS metrics (as seen in table 5). Thus, two additional variations of height-for-age and weight-for-age z-scores were created using the newly created age variable in an attempt to eliminate any remaining effect from age heaping.

First, since it was shown from the incorrect age calculation that age appears to often be rounded down to 12 or 24 months in reporting instead of providing an exact estimation of age in months, the values of 12 and 24 months are suspect to rounding. If these two values were eliminated, it would theoretically leave only exact (and un-rounded) age values. New variables were created for height-for-age z-scores and weight-for-age z-scores, and subsequently stunting and underweight prevalence variables, in which the values for 12 and 24 were eliminated.

Second, an additional variation on each z-score was created in which the age value itself is changed. Within each age band, the age of the child was changed to reflect the mid-point of the age band (e.g. 6 months for those between 0 and 11 months of age; 18 months for those between 12 and 23 months of age; and 30 for those between 24 and 35 months of age), and the respective z-score was re-calculated based upon existing height or weight values and child's sex, taking account of the newly assigned age value. By setting each child's age to the mid-point of their age band, the newly created variables eliminated any effect of heaping from mis-reporting of age.

In sum, three possible anthropometric outcomes were created based upon manipulation of child age:

Outcome 1: Z-score and prevalence estimate from newly created age variable
Outcome 2: Z-score and prevalence estimate from newly created age variable, with ages 12 and 24 removed
Outcome 3: Z-score and prevalence estimate from newly created age variable with age reset to mid-point of age band

All three outcome measures for stunting and underweight were explored in both tranches to determine if Outcome 2 or Outcome 3 changed mean values of stunting and underweight, thus eliminating any remaining effects from age heaping found in Outcome 1.

Little, if any, difference was found between Outcome 1 and 2 for either stunting or underweight estimates selecting out for either under-3 or under-2 children. Outcome 3 resulted in much higher mean z-score and prevalence estimates on nearly all accounts (often by nearly 10 ppts). Additionally, the change seen between baseline and midline estimates in Outcome 3 was much smaller than the changes seen in Outcome 1 and 2. Therefore, Outcome 3 was discontinued as it appeared to not accurately reflect changes in mean z-score and prevalence identified using more traditionally calculated outcome variables. Since the creation of Outcome 2 results in a smaller sample size after eliminating 12 and 24 values, and since there was little difference between Outcome 1 and 2, Outcome 1 was selected as the most appropriate variable to use for analysis.

BIBLIOGRAPHY

ACC/SCN (1990). *Appropriate uses of anthropometric indices in children*. Nutrition policy discussion paper No. 7. Geneva: United Nations Administrative Committee on Coordination – Subcommittee on Nutrition.

ACC/SCN (1996). *How nutrition improves*. Nutrition policy discussion paper No. 15. Geneva: United Nations Administrative Committee on Coordination – Subcommittee on Nutrition.

Alderman H, Ndiaye B, Linnemayr S, Ka A, Rokx C, Dieng K, Mulder-Sibanda M (2009). Effectiveness of a community-based intervention to improve nutrition in young children in Senegal: a difference in difference analysis. *Public Health Nutrition* 12(5):667-73.

Ballard T, Coates J, Swindale A, Deitchler M (2011). *Household Hunger Scale: Indicator Definition and Measurement Guide.* Washington, DC: FANTA-2 Bridge, FHI 360.

Bhutta ZA, Ahmed T, Black RE, Cousens S, Dewey K, Giugliani E, Haider BA, Kirkwood B, Morris SS, Sachdev HP, Shekar M for the Maternal and Child Undernutrition Study Group (2008). What works? Interventions for maternal and child undernutrition and survival. *Lancet* 371(9610):417-40.

Black RE, Allen LH, Bhutta ZA, Caulfield LE, de Onis M, Ezzati M, Mathers C, Rivera J for the Maternal and Child Undernutrition Study Group (2008). Maternal and child undernutrition: global and regional exposures and health consequences. *Lancet* 371(9608):243-60.

Brown CA, Lilford RJ (2006). The stepped wedge trial design: a systematic review. *BMC Med Res Methodol* 6:54.

Bryce J, Coitinho D, Darnton-Hill I, Pelletier D, Pinstrup-Anderson P for the Maternal and Child Undernutrition Study Group (2008). Maternal and child undernutrition: effective action at national level. *Lancet* 371(9611):510-26.

Buback L, Mason J (2011). *Addendum: Analysis of process and underweight data from CBN-ENCU in Ethiopia*. Report for UNICEF: http://tulane.edu/publichealth/internut/

Chotard S, Mason JB, Oliphant NP, Mebrahtu S, Hailey P (2010). Identifying priorities for emergency intervention from child wasting and mortality estimates in vulnerable areas of the Horn of Africa. F*ood Nutr Bull* 31(3 Suppl):S234-47.

Cook T, Campbell D (1979) *Quasi-experimental design and analysis issues for field settings*. Chicago: Rand McNally.

CSA, ICF International (2012). *Ethiopian Demographic and Health Survey 2011.* Addis Ababa & Calverton, Maryland: Central Statistical Agency and ICF International.

FAO (December 2011). *Crops and Food Shortages. No. 4. December 2011*. Rome: Food and Agriculture Organization.

FMOH (2008). *Program Implementation Manual of National Nutrition Program (NNP) – I July 2008 – June 2013.* Addis Ababa: Ethiopian Federal Ministry of Health.

Galasso E, Yua J (2006). *Learning through monitoring: lessons from a large scale nutrition program in Madagascar*. Washington, D.C: World Bank (Policy Research Working Paper 4058).

Galasso E, Umapathi N (2007). *Improving nutritional status through behavioral change: lessons from Madagascar*. Washington, D.C: World Bank (Policy Research Working Paper 4424).

Habicht JP, Victora CG, Vaughan JP (1999). Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact. *Int J Epidemiol* 28 (1):10-8.

Haines A, Sanders D, Lehmann U, Rowe AK, Lawn JE, Jan S, Walker DG, Bhutta Z (2007). Achieving child survival goals: potential contribution of community health workers. *Lancet* 369 (9579):2121-31.

Hoblitt, A, Mason J (2010). *Analysis of process and underweight data from CBN-ENCU data (first tranche of CBN, 39 woredas, mid 2008-May 2010*. Report for UNICEF: http://tulane.edu/publichealth/internut/

Mason JB (1996). Within a decade no child with go to bed hunger? *Proceedings of the Nutrition Society* 55:621-640.

Mason JB, Sanders D, Musgrove P, Soekirman, Galloway R (2006). *Community Health and Nutrition Programs.* In: Jamison DT, Breman JG, Measham AR, Alleyne G, Claeson M, Evans DB, Jha P, Mills A, Musgrove P, eds. *Disease Control Priorities in Developing Countries, 2nd edition*. Washington, DC: World Bank.

Mason JB (2008). *NNP Evaluation Plan.* 30 Sept. 2008.

Mason JB, Chotard S, Cercone E, Dieterich M, Oliphant NP, Mebrahtu S, Hailey P (2010). Identifying priorities for emergency intervention from child wasting and mortality estimates in vulnerable areas of the Horn of Africa. *Food Nutr Bull* 31(3 Suppl):S234-47.

Moriss SS, Cogill B, Uauy R for the Maternal and Child Undernutrition Study Group (2008). Effective international action against undernutrition: why has it proven so difficult and what can be done to accelerate progress? *Lancet* 371(9612):608-21.

UNSCN (2010). *6th Report on the World Nutrition Situation: Progress in nutrition*. Geneva: United Nations Subcommittee on Nutrition.

USAID (2010). *An Atlas of Ethiopian Livelihoods*. Addis Ababa: Livelihood Integration Unit.

USAID (2011). *Essential Nutrition Action Frameworks: Booklet on Key ENA Messages*.

Victora CG, Habicht JP, Bryce J (2004). Evidence-based public health: moving beyond randomized trails. *Am J Public Health* 94 (3):400-5.

Victora CG, Adair L, Fall C, Hallal PC, Martorell R, Richter L, Sachdev HS for the Maternal and Child Undernutrition Study Group (2008). Maternal and child undernutriton: consequences for adult health and human capital. *Lancet* 371(9609):340-57.

White J, Mason J (2011). *Mid-term evaluation of the CBN component of the Ethiopian NNP: Highlights of recent findings.* Report for UNICEF: http://tulane.edu/publichealth/internut/

WHO (2010a). *Indicators for assessing infant and young child feeding practices: Part 2 measurement.* Geneva: World Health Organization.

WHO (2010b). *Indicators for assessing infant and young child feeding practices: Part 1 definitions.* Geneva: World Health Organization.

WHO (2012). *Essential Nutrition Actions: Improving maternal-newborn-infant and young child health and nutrition.* Geneva: World Health Organization.

World Bank (2001). *Implementation Completion Report: Community Nutrition Project, Senegal.* Washington, D.C: The World Bank (Report No. 21429).

World Bank (2008). *Project Appraisal Document on a Proposed Grant in the amount of SDR 18.8 million to the Federal Democratic Republic of Ethiopia for a Nutrition Project*. Report No. 42171-ET. Washington, DC: The World Bank.

World Bank (2010). *What can we learn from Nutrition Impact Evaluations?* Washington, DC: The World Bank.

Yang H, de Onis M (2008). Algorithms for converting estimates of child malnutrition on the NCHS reference into estimates based on the WHO Child Growth Standards. *BMC Pediatr* 8:19.