USAID
FROM THE AMERICAN PEOPLE

COMPARATIVE COST-EFFECTIVENESS OF FOUR SUPPLEMENTARY FOODS IN TREATING MODERATE ACUTE MALNUTRITION IN CHILDREN 6-59 MONTHS IN SIERRA LEONE

A Report from the Food Aid Quality Review

Section 1: Effectiveness Study
Section 2: Body Composition Sub-Study
Section 3: Environmental Enteric Dysfunction Sub-Study
Section 4: Neurocognitive Sub-Study

This report was made possible by the generous support of the American people through the support of the United States Agency for International Development's Bureau for Humanitarian Assistance (USAID/BHA) and the legacy Office of Food for Peace (FFP) under the terms of Contract AID-OAA-C-16-00020, managed by Tufts University.

EXECUTIVE SUMMARY

In 2011, the United States Agency for International Development (USAID) Food Aid Quality Review (FAQR), led by the Friedman School of Nutrition Science and Policy at Tufts University, released a report entitled *Improving the Nutritional Quality of U.S. Food Aid: Recommendations for Changes to Products and Programs*.[1] That report made recommendations to improve the nutritional quality of food assistance commodities. Those recommendations included increasing the amount of fortified vegetable oil (FVO) programmed with corn soy blend plus (CSB+) to achieve an oil to CSB+ ratio of 30g:100g; adding a dairy component in the CSB+ formulation; adding ready-to-use foods (RUFs) as an option in programming; and increasing the use of cost-effectiveness analysis in decision-making. Before integrating these recommendations into USAID Office of Food for Peace (FFP) – as of 2020, FFP was integrated into the newly formed Bureau of Humanitarian Assistance (BHA) – food assistance programming, FAQR Phases II and III were tasked with evaluating the comparative effectiveness and cost-effectiveness of these recommendations in real field settings. From April 2017 to November 2018, FAQR Phase III tested these recommendations for the treatment of moderate acute malnutrition (MAM) among children 6-59 months in Sierra Leone.

The *Four Foods MAM Treatment Study* determined the comparative effectiveness and cost-effectiveness of four specialized nutritious foods (SNFs) in achieving recovery (defined as a mid-upper arm circumference (MUAC) of 12.5 cm or more) from MAM within 12 weeks. The four foods were:

- CSB+ with fortified vegetable oil (CSB+ w/oil)
- Corn soy whey blend with fortified vegetable oil (CSWB w/oil)
- Ready-to-use supplementary food (RUSF)
- Super Cereal Plus with amylase (SC+A)

During the period of the study's implementation, CSWB w/oil and SC+A were novel formulations not previously programmed for MAM treatment in other contexts. Both CSB+ and CSWB were distributed in a novel package size of 1.2 kg of flour as opposed to the standard 25 kg bag often repackaged in-country or at points of distribution. Rations were isocaloric across SNFs (~550 kcal/day on average). Children were followed for one, three, and six months after recovery to track incidence of relapse to MAM. The study also explored certain behavioral factors that may have influenced recovery.

Section 1: Effectiveness and Cost-effectiveness Study details the methods and results of the *Four Foods MAM Treatment Study* comparing the effectiveness and cost-effectiveness of the four foods. The methods of data collection included caregiver report of common morbidities and anthropometric measurements; in-depth quantitative surveys of a subsample exploring cooking practices, perceptions of malnutrition, and certain household behaviors or practices; five-day consecutive in-home observations on a subsample; focus group discussions; facility-based observations; and key informant interviews. Together the data were compiled to explore questions about factors that influence effectiveness and cost-effectiveness of the MAM treatment foods in achieving recovery.

In total, 2,691 children were enrolled from April 2017 to September 2018. After exclusion criteria were applied, 2,653 children remained in the sample. The unadjusted recovery rates ranged from 62.1 percent to 64.5 percent. After adjusting for covariates, there were neither significant nor observable differences in recovery among the foods. Cost per recovered child ranged from $90 to $94 (in United States dollars, USD), and cost-effectiveness analysis showed no discernable differences among the four foods. Though opportunity costs to caregivers were lowest in the RUSF arm (~$5), the difference was not enough to change the conclusion that the four foods were similarly cost-effective when combined with the program perspective. Sustained recovery at four-weeks after discharge was different between RUSF

and the comparator CSB+ (73% compared to 81%, $p<0.05$) but did not alter cost-effectiveness conclusions in cost per child who sustained recovery from the combined perspective. Intrahousehold sharing rates were similar across arms (~25%) and unassociated with recovery. Rationales for sharing were reportedly community views of the foods as being healthy and necessary for childhood well-being. Preparing or feeding the foods according to recommendations was also not associated with recovery. However, observed consumption of the food by the target child was associated with recovery.

Section 2: Body Composition, defined as the proportion of fat mass (FM) and fat-free mass (FFM) in the body, body composition has been shown to be an important indicator of health and survival in young children.[2] While FM represents energy stores, FFM represents functional organs and tissues. Inadequacy of either FM or FFM has been shown to be detrimental to health, while excess of FM has also been linked to incidence of noncommunicable diseases later in life. This sub-study assessed if any of the four SNFs differentially accreted FFM compared to FM as measured by a deuterium dilution, if differences in accretion influenced recovery, and whether changes in anthropometric measures correlated with changes in body composition.

In total, 578 children were enrolled of whom 407 completed both the baseline and 4-week measurements. Baseline body composition measures were comparable across arms at baseline (~0.26% FM ± 8.44). On average, weight gain over four weeks ranged from -0.65 kg to 2.1 kg, change in FFM ranged from -1.5 kg to 2.2 kg, and change in FM ranged from -1.8 kg to 1.6 kg. Approximately 82% of weight gained during the 4-week period was FFM. Though correlations between body composition and anthropometric measures were significant ($p<.05$) they were weak (MUAC: Pearson~0.23-0.26; weight-for-height Z-score (WHZ): Pearson ~0.24-0.25). Children who recovered within 12 weeks gained an average of 0.43 kg FFM during the first 4 weeks of treatment, while those who did not recover by 12 weeks gained an average of 0.22 kg FFM ($P = 0.002$). There were no food-specific differences in body composition measured in the first four weeks of treatment.

Section 3: Environmental Enteric Dysfunction (EED) is a disorder of the small intestine that may play a role in the etiology and treatment of MAM.[3] This sub-study assessed whether the presence of EED modified the effect of the four SNFs on growth measured by anthropometric indicators and, secondarily, whether any of these SNFs led to change in EED over four weeks of treatment. The study also assessed the association between EED and household water, sanitation, and hygiene (WASH) practices and compared the microbiota profile of MAM children with varying levels of EED. The biomarkers of EED were the lactulose to mannitol (L:M) test, 15 fecal host mRNA transcripts, and 3 fecal host proteins. The 15 mRNA transcripts were grouped using principal components analysis (PCA) into thee scores representing different components of EED: Gut Inflammation Score, Gut Permeability Score, and Gut Defense Score.

Presence of EED at enrollment using any of the biomarkers did not modify the effect of the study foods except for Gut Defense Score (GDS) ($p=0.001$). More children with high GDS recovered compared to children with lower GDS ($p<0.001$). There was no change in the fecal host mRNA transcripts between two time points or by study food. Among the biomarkers, only Gut Inflammation Score (GIS) and EED Protein Score were weakly but significantly correlated ($r=0.22$, $p<0.05$) with each other. None of the biomarkers could predict presence of EED (%L≥0.2) with high sensitivity or specificity. Eight fecal host mRNA transcripts (AQP9, REG3A, IFI30, DECR1, BIRC3, SELL, PIK3AP1, DEFA6) identified EED (%L≥0.2) and severe EED (%L≥0.45) with 85 percent sensitivity and 80 percent specificity. Differences were observed in household WASH behaviors between study participants with and without EED. These behaviors included children putting soil or animal feces in the mouth, animals observed drinking from household drinking water, and caregivers covering household drinking water storage containers with a

lid. Participants with high GIS (more inflammation) had lower bacterial diversity compared to low or medium GIS (p=0.005), and different bacterial communities were present at varying levels of GIS (p=0.009).

Section 4: Neurocognitive Function examined the feasibility, measurement properties, and utility of a novel eye-tracking test in the assessment of neurocognitive function in young children in low-resource settings. The sub-study tested whether neurocognitive processing, measured by saccadic reaction time, is measurably slower in children with MAM and whether cognitive delays may be restored to the level of non-MAM children after four weeks of treatment with supplementary food.

Data on visual orienting were obtained for 76% of the infants at the first attempt of eye tracking. Visual orienting responses to stimuli on a computer screen were frequent in MAM and control children (mean=95%), yielding sufficient data for reliable estimation of Saccadic Reaction Time (SRT) (test-retest r=0.60, P<0.001) and the mean duration of visual fixations (test-retest r=0.53, P<0.001). MAM children had marginally slower SRTs than controls in the baseline assessment. After four weeks of supplementary feeding, SRT improved among the children with MAM (-6 ms ± 2, P=0.006), as did fixation duration (+34 ms ± 6, P<0.0001). For control children, there were no significant changes in SRT (+4ms ± 3, P=0.19) or fixation duration (+8ms ± 9, P=0.43). No group or treatment effects were found in the conventional observational tests of visual orienting.

What We Learned from This Study: Main Conclusions and Recommendations

1. **Because there were no differences in effectiveness or cost-effectiveness, selection of a treatment food should be context- or program-specific, based on cost and feasibility.**

1.1 Recovery rates across the four foods were similar. This finding is consistent with results from previous studies in which neither RUFs nor any of the fortified blended foods (FBFs) performed systematically better than the others in field trials.[4,5] This suggests that when choosing which commodity to distribute for MAM treatment, decision-makers should focus on improvements to program designs, foods that are fit for purpose, and budgetary and programmatic feasibility.

1.2 The four foods were also similarly cost-effective. There are few studies exploring the comparative cost-effectiveness of different SNFs for MAM treatment, and they have reported different results.[6,7] The findings in this study found no discernable difference in the cost-effectiveness of these four foods; however, further research in this area is needed to understand which local factors should be considered in selection of SNFs.

1.3 Caregivers' opportunity costs differed by food, but this did not alter the comparative cost-effectiveness. Reducing caregiver opportunity costs has been cited as a chief factor in improving recovery rates for severe acute malnutrition (SAM).[8] Caregiver opportunity costs have also been cited as a rationale for increasing use of RUFs instead of FBFs for MAM treatment.[9,10] The results of this study do not support these conclusions in the context of MAM treatment. A joint statement on MAM treatment from the World Health Organization (WHO), United Nations Children's Fund (UNICEF), World Food Programme (WFP), and the United Nations High Commission for Refugees (UNHCR) cited a need for improved understanding of opportunity costs to inform decisions about programs and commodities.[11] The findings from this study support the recommendation for further research to identify ways in which opportunity costs influence MAM treatment and recovery.

1.4 Changes in body composition did not differ by the type of study food. As with the effectiveness results, this too is consistent with other studies which showed that over the course of supplementation children gained weight, but there were no food-specific differences in FM or FFM accretion.[12,13] Though further research into field-friendly methods of measuring body composition are warranted, there is little evidence that there are food-specific effects of SNFs on body composition.

1.5 Recovery among children with EED did not differ by study food. Presently, we are unaware of any studies exploring improvements in EED that are attributable to differences in SNF composition. Independent of food, the presence of EED — across measures — did reduce the rate of recovery from MAM. This supports the recommendation for increased investment in implementation research and evaluation of programs designed to reduce the impact of underlying biological conditions in order to improve rates of recovery.

1.6 Expanding definitions of "recovery" may change cost-effectiveness conclusions. Instead of cost per child who recovered, substituting cost per child who *sustained recovery* yielded different point estimates that rendered RUSF less cost-effective than the FBFs. We are unaware of any published studies that have examined sustained recovery through a cost-effectiveness lens. Since children who relapse into MAM after recovery require another course of treatment, sustained recovery is critical to cost-effective programming. More attention is needed to understand and appropriately measure sustained recovery.

In summary, there were no differences in recovery rates. The same was true of cost-effectiveness measures from the program perspective; the four foods were equally cost-effective measured in cost per recovered child. Caregivers' opportunity costs differed by food, but not so drastically as to alter the relative cost-effectiveness of the four foods when combined with the program perspective. Consistent with the effectiveness and cost-effectiveness findings, there were no differences in change in body composition by food. There were also no food-specific differences in recovery rates across the foods among children with EED, though the presence of EED did reduce recovery rates. These findings underscore that FFP can choose from a range of commodities with evidence supporting their use and given a range of price points. There is no single preferred product for all situations.

2. Focus on program implementation, including delivery modality and behavioral determinants of program effectiveness and cost-effectiveness.

2.1 Among the determinants of program costs, the mobile- Supplementary Feeding Program's (SFP) operational costs were the largest cost component. This contrasts with conclusions from an SFP in Mali, where the SNF was delivered in the context of an ongoing SFP, and the commodities were the largest cost contributors.[14] Unfortunately, as reported by Kennedy et. al., there continue to be few cost-effectiveness studies of SFPs generally and even fewer that specifically report details on *program costs*.[15] We need better data and evaluations of SFPs to understand what elements of program design – both food and non-food – affect effectiveness and cost-effectiveness. Investments in improving the data quality and design of cost-effectiveness studies will make it easier for researchers to compare these results to other SFP models.

2.2 Smaller sized packaging for FBFs was more expensive than conventional bagging but likely more efficient for delivery. Though there have been few formal studies examining the influence of different packaging types on the relative effectiveness of different SNFs, a recent review detailed the diverse challenges specific to SNF packaging.[16] The authors highlighted challenges with 25 kg multiwall paper bags typically used for CSB+ and differences in the box sizes for the smaller, 1.25 kg

bags normal for SC+. Because the Four Foods MAM Treatment study did not use 25 kg bags of CSB+, we cannot confirm that smaller bags mitigated the challenges reported in the review. However, a change in box size, for example, was a challenge experienced at the storage warehouse when the study switched suppliers during the second shipment. As the review recommended, harmonization of packaging and further research on the influence of packaging on programming and effectiveness should be conducted.

2.3 Procuring commodities locally was not cheaper than international tenders. Improving and expanding local procurement channels for SNFs has been urged in order to reduce costs, increase supply, and speed delivery time in emergency situations. [17] Further, research has shown that locally procured SNFs are as effective as internationally procured commodities, depending on the context and food.[18,19] However, we did not find local procurement to be less costly than international tender suggesting that policy makers should consider how to improve cost-effectiveness of local, regional, or global producers for such improvements to be cost-effective. Areas of further research might include: measuring positive externalities to local population (e.g. improvements in employment); process evaluations for local supply chains; the role of transportation; and evaluation of local quality control labs and procedures.

2.4 Program implementation applies not only to food delivery, but also to delivery modality and to complementary services such as counseling, promotion of good WASH practices, and positive behavior change. The development of the integrated community case management (ICCM) approach to treating acute malnutrition (SAM and MAM) focused attention on *delivery* as a necessary component in the effectiveness of treatment.[20] As this study has shown, the key elements determining effectiveness relate to how the food is delivered and how it is used by the family. Studies such as WASHPlus and WASHBenefits highlighted the significant association between infant and young child feeding (IYCF) practices and WHZ (i.e. wasting). These studies also highlighted the inconclusive conclusions about the relationship between WASH practices and change in WHZ or HAZ (i.e. stunting).[20,21] The findings contained in this report highlight the need for more research on delivery modalities and nutrition-sensitive components (e.g. IYCF practices) as potential pathways for sustaining recovery and/or limiting the potential for MAM ever manifesting in young children in low-resource settings.[22]

In summary, in the costing analysis, the mobile SFP's operational costs, not food-specific costs, were the largest contributors to the overall program cost. Smaller sized packages for FBFs were slightly more expensive than the traditional 25kg bags but were also more efficient for distribution and storage. Additional services such as counseling, promotion of good WASH practices, and positive behavior change were shown to be important in encouraging proper feeding and household behaviors associated with recovery. Decisions about SFP implementation should therefore consider all program components, not only the supplementary food. Implementation research is needed to inform program decision making.

3. Definitions of recovery should include indicators apart from reaching an anthropometric target

3.1 While recovery by 12 weeks was similar in all four arms, sustained recovery at four weeks after discharge was lower in the RUSF arm than in the FBFs. The overall rate of sustained recovery for all four foods was, however, similar to those reported elsewhere.[23,24,25] More information is needed detailing how diverse contexts, different MUAC cutoffs for identifying graduation,

fixed treatment protocols (e.g. a fixed 12-week treatment protocol vs. graduation immediately upon achieving a satisfactory MUAC), and longer treatment periods after recovery may influence relapse.

3.2 Children who sustained recovery tended to gain more FFM, FM, and weight over the first 4 weeks of treatment. Children gained mostly FFM (versus FM) on average, but these differences were not significant. However, children who recovered from MAM by 12 weeks were already showing differential changes in overall weight gain at 4 weeks (higher gains in both fat and fat-free mass), and the same was true of children who sustained recovery for one month after graduation. Overall, there is limited information on the changes in body composition during MAM and MAM treatment.[19] The optimum proportion of lean mass to fat mass in children recovering from MAM is also unknown and warrants further exploration.

3.3 Changes in body composition were poorly correlated with anthropometric measures. There is recognition by the international nutrition community that SFPs need to assess outcomes other than anthropometric indicators such as MUAC and WHZ.[19] The fact that alternative measures such as body composition show weak correlations with anthropometric measures supports giving more consideration to alternative methods for identifying acutely malnourished cases and for cutoffs to be used when declaring them recovered.

3.4 Cognitive performance is lower in malnourished children, but malnourished children reached the level of their well-nourished peers after just four weeks of treatment. The role of improved nutrition in improving cognitive performance is widely recognized.[26,27] This study used a novel method for measuring cognitive performance based on eye movements that is less sensitive to culture and context in interpretation than other methods and was able to measure improvements in neurological function associated with cognitive performance over four weeks of treatment.

3.5 Presence of EED reduced the effectiveness of supplemental feeding. Children with MAM who started the program with a healthier small intestine (based on GDS) were more likely to graduate from the treatment program within 12 weeks.

In summary, changes in body composition were poorly correlated with anthropometric measures. However, changes in body composition in the first four weeks of treatment were significantly different when comparing children who recovered to children who did not. Additionally, children who sustained recovery tended to gain more FFM, FM, and weight in the first four weeks of treatment. Supplementary feeding was also correlated with improvements in EED during the first four weeks of treatment. In the comparative effectiveness analysis, sustained recovery at one month was lowest in the RUSF arm compared to CSB+ w/oil. Examining cost-effectiveness through a sustained recovery lens altered results such that cost per child who sustained recovery was highest in the RUSF arm compared to CSB+ w/oil (albeit not significantly). Cognitive performance was marginally slower in malnourished children compared to their well-nourished peers, although cognitive performance of malnourished children was shown to attain the level of well-nourished peers after four weeks of treatment. These consistent results illuminate the need for more nuanced measures of recovery than traditional MUAC, height, and weight cutoffs. This also highlights the potential importance of the first four weeks of treatment in predicting long-term outcomes of supplementary feeding.

4. Empirical observations of household behavior shed light on factors associated with program effectiveness and cost-effectiveness.

4.1 Intrahousehold sharing of the foods with those other than the intended recipient was both reported and observed for all foods, but sharing was not associated with recovery. Previous studies argued that local perceptions of RUSF as a medicine led to a lower rate of sharing compared to FBFs.[28] In some studies, higher sharing rates of FBFs was presumed to be a consequence of higher levels of acceptability among the beneficiary children receiving the FBF.[29] This study found that sharing, whether measured by self-report or by direct observation, was equally likely among all four foods However the study also found that sharing was not associated with a lower likelihood of recovery.

4.2 Adherence to the recommended recipe was not associated with the likelihood of recovery. Caregivers received instructions on the preparation of the FBFs that included the correct ratio of water to flour and (for CSB+ and CSWB) the ratio of oil to flour. Adherence to these instructions was assessed by caregiver self-report and by laboratory analysis of the prepared porridge to determine fat content. There was variation in the degree of adherence to the recipe, but these variations were not associated with the likelihood of recovery. This suggests (and was confirmed by our observation) that the target child's consumption of the food supplement is key to recovery.

4.3 Children who were observed to consume the supplementary food were more likely to recover. In this study, we assessed consumption by caregiver retrospective self-report and by five-day consecutive in-home observations Children who were observed to consume the food during the observation period were more likely to recover within the 12 weeks of treatment

In summary, sharing of the treatment food was observed equally among all foods but was not a predictor of recovery. Caregivers reportedly following the recommended recipe for the study foods also was not a predictor of recovery. Consumption of the study food by the beneficiary child, however, was associated with a higher likelihood of recovery, illustrating the importance of actual consumption on improvement in MAM regardless of other behaviors.

---

[1] Webb P, Lorge Rogers B, Rosenberg I, et al. *Improving the Nutritional Quality of U.S. Food Aid: Recommendations for Changes to Products and Programs*.; 2011. https://pdf.usaid.gov/pdf_docs/pnadz841.pdf. Accessed August 14, 2018.

[2] Radhakrishna KV, Kulkarni B, Balakrishna N, Rajkumar H, Omkar C, Shatrugna V. Composition of weight gain during nutrition rehabilitation of severely under nourished children in a hospital based study from India. *Asia Pac J Clin Nutr*. 2010;19(1):8-13. doi:10.6133/apjcn.2010.19.1.02

[3] Campbell RK, Schulze KJ, Shaikh S, et al. Environmental enteric dysfunction and systemic inflammation predict reduced weight but not length gain in rural Bangladeshi children. *Br J Nutr*. 2018;119(4):407-414. doi:10.1017/S0007114517003683

[4] Annan RA, Webb P, Brown R. Management of Moderate Acute Malnutrition (MAM): Current Knowledge and Practice Collaborating to improve the management of acute malnutrition worldwide. 2014;(September). www.cmamforum.org.

[5] Lazzerini M, Rubert L, Pani P. Specially formulated foods for treating children with moderate acute malnutrition in low- and middle-income countries (Review). 2013;(6). doi:10.1002/14651858.CD009584.pub2.www.cochranelibrary.com

[6] Isanaka S, Barnhart D A, McDonald CM, et al. Cost-effectiveness of community-based screening and treatment of moderate acute malnutrition in Mali. *BMJ Glob Heal*. 2019;4(2):1-10. doi:10.1136/bmjgh-2018-001227

[7] Suri DJ, Moorthy D, Rosenberg IH. The Role of Dairy in Effectiveness and Cost of Treatment of Children with Moderate Acute Malnutrition. *Food Nutr Bull*. 2016;37(2):176-185. doi:10.1177/0379572116633327

[8] Puett C, Sadler K, Alderman H, Coates J, Fiedler JL, Myatt M. Cost-effectiveness of the community-based management of severe acute malnutrition by community health workers in southern Bangladesh. *Health Policy Plan*. 2013;28(4):386-399. doi:10.1093/heapol/czs070
[9] Nackers F, Broillet F, Oumarou D, et al. Effectiveness of ready-to-use therapeutic food compared to a corn/soy-blend-based pre-mix for the treatment of childhood moderate acute malnutrition in Niger. *J Trop Pediatr*. 2010;56(6):407-413. doi:10.1093/tropej/fmq019
[10] Lagrone L, Cole S, Schondelmeyer A, Maleta K, Manary MJ. Locally produced ready-to-use supplementary food is an effective treatment of moderate acute malnutrition in an operational setting. *Ann Trop Paediatr*. 2010;30(2):103-108. doi:10.1179/146532810X12703901870651
[11] WHO. WHO, UNICEF, WFP and UNHCR Consultation on the Programmatic Aspects of the Management of Moderate Acute Malnutrition in Children under five years of age. 2010;(February):1-18.
[12] Fabiansen C, Yameogo CW, Iuel-Brockdorf AS, et al. Effectiveness of food supplements in increasing fat-free tissue accretion in children with moderate acute malnutrition: A randomised 2 x 2 x 3 factorial trial in Burkina Faso. *PLoS Med*. 2017;14(9):e1002387. doi:10.1371/journal.pmed.1002387
[13] Mcdonald CM, Ackatia-armah RS, Doumbia S, Kupka R, Duggan CP, Brown KH. Percent Fat Mass Increases with Recovery , But Does Not Vary According to Dietary Therapy in Young Malian Children Treated for Moderate Acute Malnutrition. 2019.
[14] Isanaka S, Barnhart D A, McDonald CM, et al. Cost-effectiveness of community-based screening and treatment of moderate acute malnutrition in Mali. *BMJ Glob Heal*. 2019;4(2):e001227. doi:10.1136/bmjgh-2018-001227
[15] Kennedy E, Branca F, Webb P, Bhutta Z, Brown R. Setting the scene: An overview of issues related to policies and programs for moderate and severe acute malnutrition. *Food Nutr Bull*. 2015;36(1):S9-S14. doi:10.1177/15648265150361S102
[16] Roubert A, Johnson Q, Walton S, Webb P. *FOOD AID PACKAGING CHALLENGES AND OPPORTUNITIES : A REVIEW OF THE PACKAGING OF FORTIFIED VEGETABLE OIL , CORN SO Y BLEND PLUS , AND SUPER CEREAL PLUS A Report from the Food Aid Quality Review*. Boston, MA; 2018.
[17] Heirman BJ, Jenkins M, Rosenzweig J. Lessons learned from evaluations of the impact of WFP programmes on moderate acute malnutrition in the Sahel. 2019;(60):107-110.
[18] Osendarp S, Rogers B, Ryan K, et al. Ready-to-use foods for management of moderate acute malnutrition: Considerations for scaling up production and use in programs. *Food Nutr Bull*. 2015;36(1):S59-S64. doi:10.1177/15648265150361S110
[19] Wegner CW, Loechl C, Mokhtar N. Moderate acute malnutrition: Uncovering the known and unknown for more effective prevention and treatment. *Food Nutr Bull*. 2015;36:S3-S8. doi:10.1177/15648265150361S101
[20] Humphrey J, N Mbuya M, Moulton L, et al. Independent and combined effects of improved water, sanitation, and hygiene, and improved complementary feeding, on child stunting and anaemia in rural Zimbabwe: a cluster-randomised trial. *Lancet Glob Heal*. 2019;7:132-179. doi:10.1016/S2214- 109X(18)30374-7
[21] Pickering AJ, Null C, Winch PJ, et al. The W ASH Benefits and SHINE trials: interpretation of W ASH intervention effects on linear growth and diarrhoea. *Lancet Glob Heal*. 2019;7(8):e1139- e1146. doi:10.1016/S2214-109X(19)30268-2
[22] Friis H, Michaelsen KF , W ells JC. Choice of design and outcomes in trials among children with moderate acute malnutrition. *Food Nutr Bull*. 2015;36(1):S35-S40. doi:10.1177/15648265150361S106
[23] Chang CY, Trehan I, Wang RJ, et al. Children Successfully Treated for Moderate Acute Malnutrition Remain at Risk for Malnutrition and Death in the Subsequent Year after Recovery. *J Nutr*. 2013;143(2):215-220. doi:10.3945/jn.112.168047
[24] Trehan I, Banerjee S, Murray E, et al. Extending Supplementary Feeding for Children Younger Than 5 Years With Moderate Acute Malnutrition Leads to Lower Relapse Rates. *J Pediatr Gastroenterol Nutr*. 2015;60(4):544-549. doi:10.1097/MPG.0000000000000639
[25] Stobaugh HC, Bollinger LB, Adams SE, et al. Effect of a package of health and nutrition services on sustained recovery in children after moderate acute malnutrition and factors related to sustaining recovery: A cluster-randomized trial. *Am J Clin Nutr*. 2017;106(2):657-666. doi:10.3945/ajcn116.149799
[26] Black RE, Allen LH, Bhutta ZA, et al. Maternal and child undernutrition: global and regional exposures and health consequences. *Lancet*. 2008;371(9608):243-260. doi:10.1016/S0140- 6736(07)61690-0
[27] Ruel MT, Alderman H. Nutrition-sensitive interventions and programmes: How can they help to accelerate progress in improving maternal and child nutrition? *Lancet*. 2013;382(9891):536-551. doi:10.1016/S0140-6736(13)60843-0
[28] Karakochuk CD, van den Briel T, Stephens D, Zlotkin S. Food Sharing Practices in Households Receiving Supplemental Foods for the Treatment of Moderate Acute Malnutrition in Ethiopian Children. *J Hunger Environ Nutr*. 2015;10(3):343-355. doi:10.1080/19320248.2014.962772
[29] Matilsky DK, Maleta K, Castleman T, Manary MJ. Supplementary Feeding with Fortified Spreads Results in Higher Recovery Rates Than with a Corn/Soy Blend in Moderately Wasted Children. J Nutr. 2009;139(4):773-778. doi:10.3945/jn.108.104018

USAID
FROM THE AMERICAN PEOPLE

Comparative Cost-Effectiveness of Four Supplementary Foods in Treating Moderate Acute Malnutrition in Children 6-59 Months in Sierra Leone

Section 1: Effectiveness Study

A Report from the Food Aid Quality Review

Prepared by:

Stacy Griswold
Breanne Langlois
Ye Shen
Ilana Cliffer
Akriti Singh
Isabel Potani
Katelyn Riffenburg
Kaiti Chatterton
Jukka Leppänen
Devika Suri
Kenneth Chui
Stephen Vosti
Shelley Walton
Lindsey Green
Memuna Sawi
Aminata Koroma
Donna Wegner
Mark J. Manary
Irwin Rosenberg
Patrick Webb
Beatrice Lorge Rogers

JANUARY 2021

This report was made possible by the generous support of the American people through the support of the United States Agency for International Development's Bureau for Humanitarian Assistance (USAID/BHA) and the legacy Office of Food for Peace (FFP) under the terms of Contract AID-OAA-C-16-00020, managed by Tufts University.

The contents are the responsibility of Tufts University and its partners in the Food Aid Quality Review (FAQR) and do not necessarily reflect the views of USAID or the United States Government.

The authors have no conflict of interest to declare.

Recommended Citation

Griswold, Stacy; Langlois, Breanne; Shen, Ye; Cliffer, Ilana; Singh, Akriti; Potani, Isabel; Riffenburg, Katelyn; Chatterton, Kaiti; Leppänen, Jukka; Suri, Devika; Chui, Kenneth; Vosti, Stephen; Walton, Shelley; Green, Lindsey Ellis; Sawi, Memuna; Koroma, Aminata; Wegner, Donna; Manary, Mark J.; Rosenberg, Irwin; Webb, Patrick; Rogers, Beatrice. 2020. *Comparative Cost-effectiveness of Four Supplementary Foods in Treating Moderate Acute Malnutrition in Children 6-59 Months in Sierra Leone*, Report to USAID from the Food Aid Quality Review. Boston, MA: Tufts University.

For correspondence, contact:

Beatrice Lorge Rogers, PhD
Friedman School of Nutrition Science and Policy
Tufts University
150 Harrison Avenue
Boston, MA 02111
beatrice.rogers@tufts.edu
http://www.foodaidquality.org

ABBREVIATIONS AND ACRONYMS

BHA	Bureau for Humanitarian Assistance (USAID)
CHW	Community Health Worker
CI	Confidence Interval
CSB	Corn-Soy Blend
CSB+	Corn-Soy Blend Plus
CSWB	Corn-Soy-Whey Blend
DFN	Directorate for Food and Nutrition
DHMT	District Health Management Team
EED	Environmental Enteric Dysfunction
FAQR	Food Aid Quality Review
FBF	Fortified Blended Flour
FFM	Fat Free Mass
FFP	Office of Food for Peace (USAID)
FGD	Focus Group Discussion
FM	Fat Mass
FVO	Fortified Vegetable Oil
GDS	Gut Defense Score
HFIAS	Household Food Insecurity Access Scale
IDI	In-depth Interview
IHO	In-home Observation
IYCF	Infant and Young Child Feeding
IRB	Institutional Review Board
LNS	Lipid-based Nutrition Supplement
MAM	Moderate Acute Malnutrition
MDD	Minimum Diet Diversity
MEST	Ministry of Education, Science and Technology
MoHS	Ministry of Health and Sanitation
MT	Metric Ton
MUAC	Mid-Upper Arm Circumference
PCA	Principal Components Analysis
PHU	Peripheral Health Unit
PPB	Project Peanut Butter
RA	Research Assistants (enumerators)
RNI	Recommended Nutrient Intake
RUF	Ready-to-Use Food
RUSF	Ready-to-Use Supplementary Food
RUTF	Ready-to-Use Therapeutic Food
SAM	Severe Acute Malnutrition
SBCC	Social and Behavior Change Communication
SC+	Super Cereal Plus
SC+A	Super Cereal Plus with Amylase
SES	Socioeconomic Status
SFP	Supplementary Feeding Program
SNF	Specialized Nutritious Food
SRT	Saccadic Reaction Time
SRT	Saccadic Reaction Time
USAID	United States Agency for International Development
USD	United States Dollars ($)
USDA	United States Department of Agriculture
WASH	Water, Sanitation, and Hygiene
WashU	Washington University in St. Louis, Missouri
WFP	World Food Programme (United Nations)
WHO	World Health Organization
WHZ	Weight for Height Z-Score
WPC	Whey Protein Concentrate

TABLE OF CONTENTS

Section 1: Main Report

ABSTRACT

In 2011, the United States Agency for International Development (USAID) Food Aid Quality Review (FAQR), led by the Friedman School of Nutrition Science and Policy at Tufts University, released a report entitled *Improving the Nutritional Quality of U.S. Food Aid: Recommendations for Changes to Products and Programs.*[1] That report made recommendations to improve the nutritional quality of food assistance commodities. Those recommendations included increasing the amount of fortified vegetable oil (FVO) programmed with corn soy blend plus (CSB+) to achieve an oil to CSB+ ratio of 30g:100g; adding a dairy component in the CSB+ formulation; adding ready-to-use foods (RUFs) as an option in programming; and increasing the use of cost-effectiveness analysis in decision-making. Before integrating these recommendations into USAID Office of Food for Peace (FFP) – as of 2020, FFP was integrated into the newly formed Bureau of Humanitarian Assistance (BHA) – food assistance programming, FAQR Phases II and III were tasked with evaluating the comparative effectiveness and cost-effectiveness of these recommendations in real field settings. From April 2017 to November 2018, FAQR Phase III tested these recommendations for the treatment of moderate acute malnutrition (MAM) among children 6-59 months in Sierra Leone.

The *Four Foods MAM Treatment Study* determined the comparative effectiveness and cost-effectiveness of four specialized nutritious foods (SNFs) in achieving recovery (defined as a mid-upper arm circumference (MUAC) of 12.5 cm or more) from MAM within 12 weeks. The four foods were:

- CSB+ with fortified vegetable oil (CSB+ w/oil)
- Corn soy whey blend with fortified vegetable oil (CSWB w/oil)
- Ready-to-use supplementary food (RUSF)
- Super Cereal Plus with amylase (SC+A)

During the period of the study's implementation, CSWB w/oil and SC+A were novel formulations not previously programmed for MAM treatment in other contexts. Both CSB+ and CSWB were distributed in a novel package size of 1.2 kg of flour as opposed to the standard 25 kg bag often repackaged in-country or at points of distribution. Rations were isocaloric across SNFs (~550 kcal/day on average). Children were followed for one, three, and six months after recovery to track incidence of relapse to MAM. The study also explored certain behavioral factors that may have influenced recovery.

Section 1: Effectiveness and Cost-effectiveness Study details the methods and results of the *Four Foods MAM Treatment Study* comparing the effectiveness and cost-effectiveness of the four foods. The methods of data collection included caregiver report of common morbidities and anthropometric measurements; in-depth quantitative surveys of a subsample exploring cooking practices, perceptions of malnutrition, and certain household behaviors or practices; five-day consecutive in-home observations on a subsample; focus group discussions; facility-based observations; and key informant interviews. Together the data were compiled to explore questions about factors that influence effectiveness and cost-effectiveness of the MAM treatment foods in achieving recovery.

In total, 2,691 children were enrolled from April 2017 to September 2018. After exclusion criteria were applied, 2,653 children remained in the sample. The unadjusted recovery rates ranged from 62.1 percent to 64.5 percent. After adjusting for covariates, there were neither significant nor observable differences in recovery among the foods. Cost per recovered child ranged from $90 to $94 (in United States dollars, USD), and cost-effectiveness analysis showed no discernable differences among the four foods. Though opportunity costs to caregivers were lowest in the RUSF arm (~$5), the difference was not enough to change the conclusion that the four foods were similarly cost-effective when combined with the program perspective. Sustained recovery at four-weeks after discharge was different between RUSF and the comparator CSB+ (73% compared to 81%, $p<0.05$) but did not alter cost-effectiveness

conclusions in cost per child who sustained recovery from the combined perspective. Intrahousehold sharing rates were similar across arms (~25%) and unassociated with recovery. Rationales for sharing were reportedly community views of the foods as being healthy and necessary for childhood well-being. Preparing or feeding the foods according to recommendations was also not associated with recovery. However, observed consumption of the food by the target child was associated with recovery.

What We Learned from This Study: Main Conclusions and Recommendations:

1. **Because there were no differences in effectiveness or cost-effectiveness, selection of a treatment food should be context- or program-specific, based on cost and feasibility.**

1.1 Recovery rates across the four foods were similar. This finding is consistent with results from previous studies in which neither RUFs nor any of the fortified blended foods (FBFs) performed systematically better than the others in field trials.[4,5] This suggests that when choosing which commodity to distribute for MAM treatment, decision-makers should focus on improvements to program designs, foods that are fit for purpose, and budgetary and programmatic feasibility.

1.2 The four foods were also similarly cost-effective. There are few studies exploring the comparative cost-effectiveness of different SNFs for MAM treatment, and they have reported different results.[6,7] The findings in this study found no discernable difference in the cost-effectiveness of these four foods; however, further research in this area is needed to understand which local factors should be considered in selection of SNFs.

1.3 Caregivers' opportunity costs differed by food, but this did not alter the comparative cost-effectiveness. Reducing caregiver opportunity costs has been cited as a chief factor in improving recovery rates for severe acute malnutrition (SAM).[8] Caregiver opportunity costs have also been cited as a rationale for increasing use of RUFs instead of FBFs for MAM treatment.[9,10] The results of this study do not support these conclusions in the context of MAM treatment. A joint statement on MAM treatment from the World Health Organization (WHO), United Nations Children's Fund (UNICEF), World Food Programme (WFP), and the United Nations High Commission for Refugees (UNHCR) cited a need for improved understanding of opportunity costs to inform decisions about programs and commodities.[11] The findings from this study support the recommendation for further research to identify ways in which opportunity costs influence MAM treatment and recovery.

1.4 Changes in body composition did not differ by the type of study food. As with the effectiveness results, this too is consistent with other studies which showed that over the course of supplementation children gained weight, but there were no food-specific differences in FM or FFM accretion.[12,13] Though further research into field-friendly methods of measuring body composition are warranted, there is little evidence that there are food-specific effects of SNFs on body composition.

1.5 Recovery among children with EED did not differ by study food. Presently, we are unaware of any studies exploring improvements in EED that are attributable to differences in SNF composition. Independent of food, the presence of EED — across measures — did reduce the rate of recovery from MAM. This supports the recommendation for increased investment in implementation research and evaluation of programs designed to reduce the impact of underlying biological conditions in order to improve rates of recovery.

1.6 Expanding definitions of "recovery" may change cost-effectiveness conclusions. Instead of cost per child who recovered, substituting cost per child who *sustained recovery* yielded different point estimates that rendered RUSF less cost-effective than the FBFs. We are unaware of any published studies that have examined sustained recovery through a cost-effectiveness lens. Since children who relapse into MAM after recovery require another course of treatment, sustained recovery is critical to cost-effective programming. More attention is needed to understand and appropriately measure sustained recovery.

In summary, there were no differences in recovery rates. The same was true of cost-effectiveness measures from the program perspective; the four foods were equally cost-effective measured in cost per recovered child. Caregivers' opportunity costs differed by food, but not so drastically as to alter the relative cost-effectiveness of the four foods when combined with the program perspective. Consistent with the effectiveness and cost-effectiveness findings, there were no differences in change in body composition by food. There were also no food-specific differences in recovery rates across the foods among children with EED, though the presence of EED did reduce recovery rates. These findings underscore that FFP can choose from a range of commodities with evidence supporting their use and given a range of price points. There is no single preferred product for all situations.

Introduction: In 2011, the United States Agency for International Development (USAID) Food Aid Quality Review (FAQR), led by the Friedman School of Nutrition Science and Policy at Tufts University, released a report entitled *Improving the Nutritional Quality of U.S. Food Aid: Recommendations for Changes to Products and Programs.*[1] That report made recommendations to improve the nutritional quality of the specialized nutritious foods (SNFs) provided in food assistance programs aimed at improving nutritional status. Those recommendations included increasing the amount of fortified vegetable oil (FVO) programmed with corn soy blend plus (CSB+) to achieve an oil to CSB+ ratio of 30 g to 100 g; adding a dairy component in the CSB+ formulation; adding ready-to-use foods (RUFs) as an option in programming; and increasing the use of cost-effectiveness analysis in decision-making. From April 2017 to November 2018, FAQR Phase III tested these recommendations for the treatment of moderate acute malnutrition (MAM) among children 6-59 months in Sierra Leone.

Methods: This was a cluster-randomized trial operating through a supplementary feeding program (SFP) providing SNFs for treatment of MAM. The study foods were: Super Cereal Plus w/ amylase (SC+A), CSB+ w/oil, Corn-Soy-Whey Blend w/oil (CSWB w/oil), and Ready to Use Supplementary Food (RUSF). Children with MAM, defined as mid-upper arm circumference (MUAC) ≥11.5 cm and <12.5 cm without bipedal edema, were enrolled at participating health clinics and received rations biweekly until they reached an outcome with a maximum of 7 rations received (approximately 12 weeks of treatment). An activity-based ingredients approach was used to measure costs per child enrolled and cost per child recovered in 2018 USD. Recovered children were followed for one, three, and six months after recovery to track incidence of relapse to MAM. A subset of participants was observed in the home for five consecutive days and/or interviewed with an in-depth survey or participated in focus group discussions.

Results: 2,683 children were enrolled out of a planned sample size of ~5,000. Overall, 63% graduated from SFP, 19% developed severe acute malnutrition (SAM), 7% defaulted (missed three consecutive visits), 1% died, and 10% remained MAM within 12 weeks. Twenty-five percent were transferred into the study from SAM treatment once they reached MAM status. By study arm, graduation rates were: 62% in CSWB w/oil, 65% in SC+A, 64% in CSB+ w/oil, 62% in RUSF. In adjusted and unadjusted models, no statistically significant differences in graduation rates among the arms were detected. A difference was observed in adjusted and unadjusted models of relapse: more children receiving RUSF relapsed within one month than those receiving CSB+ w/oil, the reference arm. Product and international freight were greatest drivers of cost differences across arms. Cost per enrolled child ranged from $86 in RUSF to $94 in SC+A. Cost per recovered child was $137 ($130 - 145) in RUSF, $142 ($134 - 151) in CSB+ w/oil, $146 ($138 - 155) in SC+A, and $149 ($140 - 160) in CSWB w/oil. Sharing was equally prevalent in all arms. Neither sharing nor adherence to recipe directions affected the likelihood of graduation but observed consumption of the supplementary food was associated with a higher likelihood of graduation.

Conclusions: Foods should be fit for purpose to maximize the cost-effectiveness of the supplementary feeding program. Since all four foods showed similar effectiveness and cost effectiveness, other programmatic factors should drive food selection. Household behavior does play an important role in effectiveness, with more focus needed to ensure consumption of the food by the beneficiary. Though no single food was found to be more effective in achieving graduation, RUSF was less effective in achieving sustained recovery. Further research is needed to understand factors that influence relapse; these may be different from behaviors influencing recovery.

1. Introduction and Background

Title II of the United States Food for Peace Act is the primary funding mechanism through which the U.S. Congress authorizes and funds international food assistance programs. [30] The Food for Peace Title II program(s) has been the dominant U.S. food assistance program since the 1980s.[30] The goal of Title II programs is to reduce the food insecurity of vulnerable populations in low income countries. Title II in-kind donations are food commodities sourced in the United States and shipped to the recipient country for distribution to target populations. The United States Agency for International Development's Office of Food for Peace (USAID/FFP), in collaboration with the United States Department of Agriculture (USDA), is responsible for overseeing Title II in-kind food assistance programs. To understand the scale of this assistance provision, in fiscal year 2018 alone, USAID/FFP distributed food assistance to more than 70 million vulnerable people globally at an estimated cost of more than USD 3 billion.[31]

As part of continuous efforts to improve the *quality* and *efficiency* of USAID/FFP Title II food assistance programs, USAID/FFP initiated the Food Aid Quality Review (FAQR) Phase 1, implemented by the Friedman School of Nutrition Science and Policy at Tufts University (Tufts) beginning in 2009. FAQR's task was to review the "state of the evidence" related to Title II food assistance as well as the nutrition-related needs of traditionally vulnerable populations who are recipients of this assistance. The findings of the review — completed in 2011 — were detailed in *Improving the Nutritional Quality of U.S. Food Aid: Recommendations for Changes to Products and Programs*.[1] Specific recommended changes to the existing USAID/FFP food commodity list included updating the micronutrient premix and the macronutrient profiles of existing fortified blended flours (FBFs); considering the addition of a dairy component to the FBF typically programmed by USAID/FFP, – Corn Soy Blend version 13 (CSB13); including ready-to-use foods (RUFs) in the food basket available for programming; and improving the cost-effectiveness evidence base for deciding when to program which commodities, given diverse programmatic needs (i.e., allowing for "fit for purpose" programming). Since 2011, the FAQR team has tested the applicability of these recommendations in field settings in Malawi and Burkina Faso.[32,33] The study detailed in this report represents the first time the effectiveness of these recommendations was tested in field settings specifically for the treatment of moderate acute malnutrition (MAM) in children 6-59 months of age in Sierra Leone.

2. Study Rationale

Specialized nutritious food products (SNFs) have been proven to be effective in treating MAM, although there remains no globally accepted standard of care.[4] SNFs are categorized into one of two groups: FBFs or RUFs. The FBFs are flours composed of a mix of precooked and milled cereals and legumes that have been fortified with a micronutrient premix, sometimes with an animal-sourced protein; vegetable oil fortified with vitamins A and D may be added to the flour at the time of manufacture or provided separately to be added to the porridge prepared in the home. RUFs are lipid-based nutritional supplements commonly composed of peanuts, soybeans, and milk powder formulated as pastes and fortified with a micronutrient premix, often with an animal-sourced protein and additional fortified vegetable oil.[4] Different studies have compared these diverse formulations, yielding inconsistent results regarding which SNF groups or formulations may be most effective for MAM treatment.[34]

From April 2017 to November 2018, a research team from FAQR Phase III implemented a study in southern Sierra Leone examining the comparative cost-effectiveness of four specific SNFs for use in the treatment of MAM (hereafter "Four Foods Study"). Based on the recommendations of the 2011 FAQR report, the four foods tested in the Four Foods Study (see **Section 2.2** for more details) were:

- USAID/FFP FBF formulation: Corn Soy Blend Plus (also referred to as "Super Cereal") distributed with fortified vegetable oil (**CSB+ w/oil**)

- FAQR's recommended reformulation of CSB+ to include a dairy component, w/oil: Corn Soy Whey Blend with fortified vegetable oil (**CSWB w/oil**)
- A novel formulation of World Food Programme's commonly programmed FBF: Super Cereal Plus with amylase (**SC+A**), and
- The standard RUF formulation for MAM treatment: ready-to-use-supplementary food (**RUSF**)

In addition to determining the comparative cost-effectiveness of the four foods, the study sought to identify behaviors that may be associated with effectiveness, such as preparation and consumption or sharing and selling of SNFs either within or between households, whether any of these behaviors differed by food, and whether they differentially influenced recovery from MAM.

2.1 Specific Objectives

The Four Foods Study had three main objectives:

Objective 1: Effectiveness
Determine whether there are differences in recovery rates among children with MAM treated with one of four supplementary foods.

Objective 2: Cost Effectiveness
Assess program costs among the four study arms, including costs of procurement, supply chain, clinic operations with food distribution, community-based activities, and caregivers' time and out-of-pocket expenses. Use these costs to determine and compare the cost-effectiveness of each SNF, measured in (1) cost per case of MAM recovered and (2) cost per recovered case of MAM that sustained recovery for 4 weeks after graduation from SFP.

Objective 3: Factors Influencing Effectiveness
Identify intervention components, behaviors, demographics, and other factors associated with effectiveness. Specifically, the study explored factors related to biweekly ration collection, preparation, consumption, sharing, and selling behaviors; knowledge and communication of social and behavior change communication (SBCC) messages regarding biweekly ration use among health workers and beneficiary mothers, and determine if any of these were different across foods.

2.2 The Four Foods in Detail

As summarized in the introduction, the Four Foods Study in Sierra Leone tested four foods: CSB+ w/oil, CSWB w/oil, SC+A, and RUSF. The differences among these four foods are described below and with greater detail in **Appendix 1**.

A. CSB+ w/oil

CSB+ is the upgraded formulation of CSB13.[35] The new formulation of CSB13 was predicated on two factors: that the micronutrient premix and macronutrient content be equal to 115 percent of recommended nutrient intake (RNI), and that any formulation changes result in a product that is at least as cost-effective as CSB13. The improved micronutrient premix included increases in zinc and vitamins B_1, B_2, B_3, B_5, B_{12}, D_3, and E; the addition of potassium and vitamin K; and the combination of two forms of iron (NaFeEDTA and ferrous fumarate) to improve absorption. Additionally, the new premix

decreased the amounts of magnesium, calcium, iodine, sodium, and vitamin A. To improve the macronutrient profile and increase caloric density, the recommended quantity of FVO included in the biweekly ration and added to the prepared porridge was increased to a ratio of 30 g FVO for every 100g of FBF.[35]

B. CSWB w/oil

To address the recommendation to add an animal-sourced protein without excessively increasing costs of the product, the FAQR team recommended that whey protein concentrate with 80 percent protein content (WPC80) be added to CSB+ w/oil instead of the alternative, dried skim milk. In nutritional value, WPC80 is protein-dense and, in small quantities, its addition to CSB+ would not displace other nutrients already in the mix, while still adding important amino acids and growth promoting factors necessary for recovery. Though the cost per metric ton (MT) of WPC80 was higher than dry skim milk at the time of the report's publication in 2011, [1] its historical price in the United States has been less variable than that of dried skim milk, giving WPC80 an element of stability for program planning.[35] The new product, CSWB, which is programmed with oil in the same ratios as CSB+, remains novel and in need of further field testing to understand the potential effects of added WPC80 on overall effectiveness and cost-effectiveness in field settings.

C. SC+A

An alternative name for CSB+ flour is Super Cereal, and the separately branded Super Cereal has been a commonly programmed FBF for WFP food assistance programs. The alternate formulation of Super Cereal is Super Cereal Plus (SC+). SC+ is CSB+ with the FVO added into the flour at the time of manufacturing rather than provided as a separate ingredient to be mixed at the home. Since 2011, WFP has experimented with updated formulations in laboratory settings, with one iteration resulting in the addition of the amylase enzyme. Amylase reduces viscosity by breaking down carbohydrates in the cooking process to ease digestion and absorption without decreasing the energy density of FBF porridges.[36] Previous studies have shown that energy intake by young children increases when amylase is a part of the SC+ mix and that acceptability is not degraded with a thinner texture.[37] Also, a novel commodity such as CSWB w/oil or SC+A requires field testing to gauge comparative effectiveness and cost-effectiveness before being included in the WFP's standard commodity list.

D. RUSF

RUSF is a modified formulation of ready-to-use therapeutic food (RUTF) that contains proportionately less milk powder that is also cheaper to produce. It is supplied in individually packaged metalized polyethylene sachets with a 24-month shelf life and is generally made with heat-treated oil seeds, pulses, cereals, sugar, milk powder, vegetable oils, vitamins, and minerals; it requires no preparation.[10]

2.3 Comparison of the Four Foods

Table 1: General Differences Among the Four Study Foods

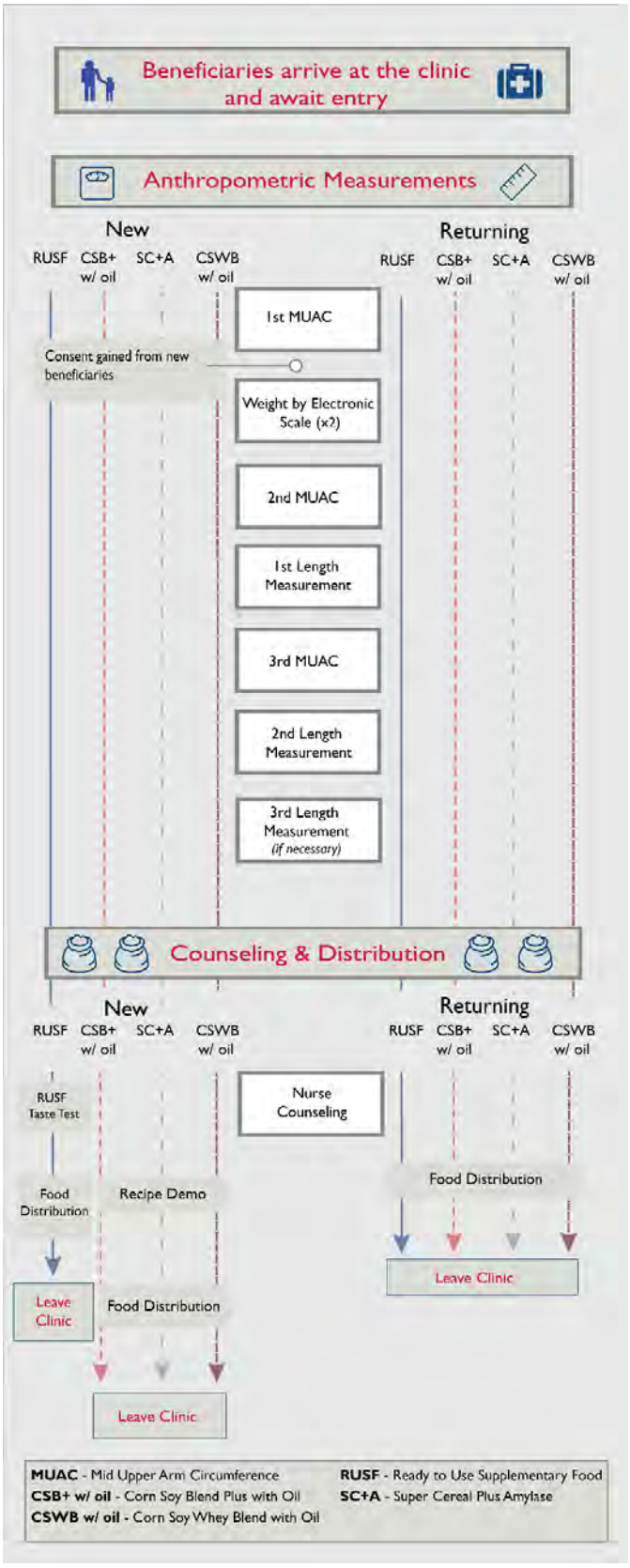
	CSB+ w/Oil	CSWB w/Oil	SC+A	RUSF
Fortified Blended Flour (FBF)	✓	✓	✓	
Programmed Separately with FVO	✓	✓		
Ready-to-Use				✓
Dairy Component (Milk or Whey)		✓	✓	✓
Experimental Formulation		✓	✓	
Experimental Packaging (1.25 kg bag)	✓	✓		

Differences among the four SNFs and their daily rations are summarized in **Table 1,** with additional detail and nutrient breakdowns in **Appendix 1**. The FBFs were distributed in hermetically sealed bags with plastic foil and a metalized layer. The packaging for the CSB+ and CSWB flours was a novel size (~1.2 kg), large enough to contain a biweekly ration (14 servings), similar in design to the 1.9 kg bags used by WFP when distributing the standard SC+ for MAM treatment. SC+A was packaged in the 1.9 kg standard size and model bags. Within the CSB+ w/oil and CSWB w/oil arms, FVO was distributed after repackaging from 4-liter metal cannisters into standard 0.5-liter plastic bottles. These foods also required preparation in the home by combining the flour with water (and oil in the case of CSB+ and CSWB) and cooking for five to ten minutes to make a porridge-like cereal. The fourth food, RUSF, required neither mixing nor cooking. The beneficiary received 14 sachets per clinic visit. If prepared and consumed by the beneficiary child as directed, all four treatment foods provided the same amount of energy (i.e., rations were isocaloric), approximately 550 kcal/day.

These packaging types contrast with what is typical for FBFs, which are distributed in either standard biweekly rations (e.g., 1.9 kg bags as with SC+) or in 25 kg bags that are repackaged at some point during the distribution chain. These differences in packaging may affect cost-effectiveness. The CSB+ and CSWB packaging used in this study (1.2 kg bags) is more expensive than the commonly programmed 25 kg bags. However, the smaller bags may reduce losses, possibly potential for contamination, and time costs from repacking that occurs along the supply chain, and may also improve the efficiency of distribution. The potential impacts of these differences on opportunity costs and packaging costs and overall cost-effectiveness are explored in more detail in the cost-effectiveness section (**Section 9**) of this report.

All the foods were distributed using a uniform clinic and distribution protocol. (See **Figure 1** for detail on SFP flow.) The only difference in clinic flow across the arms was for new beneficiaries: those receiving any of the FBFs were required to stay to the end of the clinic in order to see and participate in a cooking demonstration, whereas those in the RUSF arm received a taste test during counseling (see **Section 5.1** and **Section 5.2** for more details). On occasion, clinic staff asked caregivers to participate in the demonstration a second time if the caregiver was unable to demonstrate an understanding of the recipes during the counseling session.

Figure 1: Clinic Flow for Both New and Returning Beneficiaries



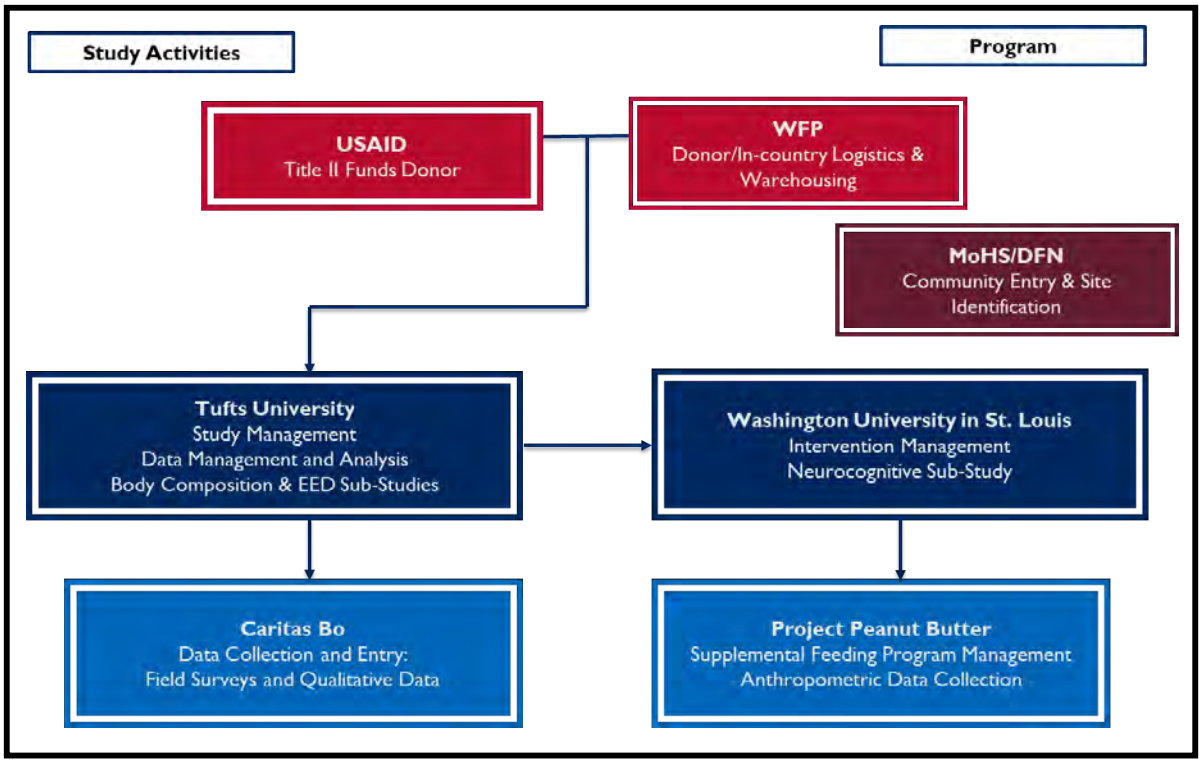
3. Partners and Institutional Roles

The Four Foods Study was made possible through a collaboration between USAID/FFP, Tufts, the School of Medicine at Washington University in St. Louis, Missouri (WashU), WFP, Sierra Leone's Directorate of Food and Nutrition in the Ministry of Health and Sanitation (DFN-MoHS), Project Peanut Butter (PPB), and Caritas Bo (**Figure 2**).

DFN-MoHS identified Pujehun District, in southern Sierra Leone, as the study site and provided technical support during the design phase of the study (see **Section 4** for more details). Within Pujehun District, the local District Health Management Team (DHMT) gave approval for the SFP to operate at specific peripheral health units (PHUs – see **Section 6** for more detail) and provided support in gaining community entry for the PPB team during start-up. Over the course of the study, the DHMT also conducted monitoring visits to strengthen collaboration between PPB and PHU clinic staff. WashU, through a partnership with PPB, managed the implementation of the SFP in Pujehun District and was responsible for screening children for MAM and collecting all clinical data from beneficiaries, including anthropometric measures and morbidity information in the previous two-week period. Tufts held primary responsibility for monitoring the technical quality of the field data collection and managing the overall study's progress. Caritas Bo gathered all field survey data, entered all data, facilitated all focus group discussions, and transcribed all focus group discussion recordings and translated them from Mende to English. WFP provided logistical support on management of the four foods, warehousing support, and in-country transport between warehouses, and managed the importation of the FBFs and the procurement of the RUSF.

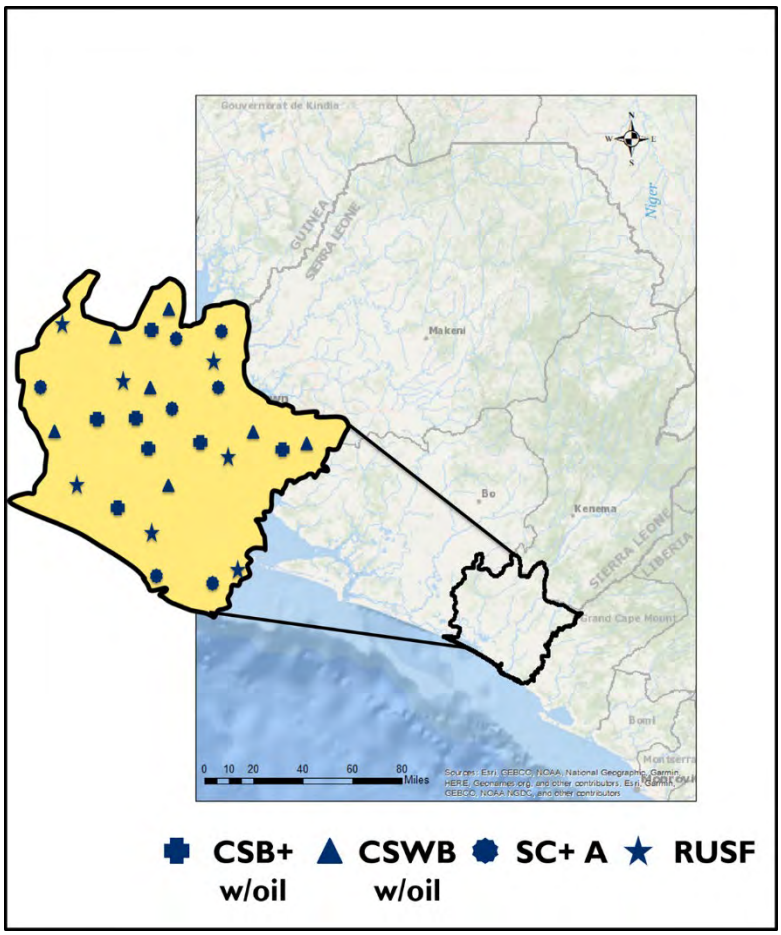
3.1 SFP Team and Mobile Clinic Model

Figure 2: Institutional Partners, Roles and Responsibilities of the Four Foods Study



The SFP was implemented by three different PPB teams, each composed of a trained nurse fluent in Mende, a clinic manager, and a driver. Under the mobile-clinic model, each day the three PPB teams loaded the appropriate quantity and type of food and necessary supplies into a Toyota Land Cruiser, which transported the team to the SFP sites and back to the PPB office in Pujehun. In this model, although government facilities (PHUs) served as the unit of randomization, the SFP was a parallel program (i.e., not integrated into MoHS service provision). **Figure 1** provides a detailed view of the SFP's data collection and daily flow from arrival to exit, and **Figure 3** provides a map of each SFP's location across Pujehun District.

Figure 3: Representative Map of Pujehun District and Study Sites



3.2 Field Data Collection Team

All enumerators (12 males and 8 females) spoke fluent Mende and had completed an undergraduate degree in social sciences. In order to gain entry to homes and be allowed to observe female caregivers' preparation behaviors, the study team determined that only females could safely observe fellow females for long periods of time. Therefore, to respect cultural sensitivities, only females collected data for the in-home observations.

To identify 20 qualified enumerators, the study team recruited 52 candidates. Over the course of six weeks, the candidates were trained in the study's methods and given rigorous ethical training and in-depth training on all the data collection instruments. The students also translated and adapted the informed consent forms. The final exam was three parts: a practical in which students conducted an enrollment interview on a role-playing caregiver, an oral Mende recitation of an informed consent form chosen at random, and a written exam covering topics from ethical scenarios to survey questions. The top scorers were hired, and the next five were placed on a roster should a backup enumerator be needed. Refresher trainings were held approximately every six months.

4. Context

Sierra Leone is a relatively small country (72,000 square km) situated on the western coast of sub-Saharan Africa, between the Republic of Guinea to the north and east and Liberia to the south.[38] Despite robust growth measured in annual change in gross domestic product(GDP) over the previous 20 years, economic growth has stagnated since the onset of the Ebola Virus Disease (EVD) crisis in 2014, and Sierra Leone remains one of the poorest countries in the world.[38,39] The country is administratively divided into four regions (Northwest, South, East, and West), subdivided into 16 districts, and subdivided further into approximately 192 chiefdoms. The climate is tropical, with the main rainy season lasting from May to November, and the dry season comprising the rest of the year. [39] The long rainy season provides conditions well suited to the production of rice, the main dietary staple. In addition to rice, the Sierra Leonean diet consists of green leafy vegetables such as cassava leaf, groundnuts/peanuts ground into paste, and palm oil, which serves as the main ingredient in sauces. [38]

In 2014, FAQR Phase II implemented a study in the Kenema District of Sierra Leone to compare the cost-effectiveness of four foods used in the treatment of MAM in children under five. Due to the Ebola Virus Disease (EVD) outbreak, the study was terminated after a short recruitment period, resulting in an insufficient sample to detect differences in effectiveness or cost-effectiveness. Understanding the comparative effectiveness and cost-effectiveness of the four foods remained a priority research endeavor under FAQR Phase III. When Sierra Leone was declared EVD-free in 2015, the FAQR team asked DFN-MoHS if Sierra Leone could be considered as a location for a new Four Foods MAM Treatment Study. DFN-MoHS was supportive of the study and identified Pujehun District as the study site.

At the time of scoping for the new study, Sierra Leone reported poor health outcomes among children under five measured in high rates of infant mortality (56 deaths/1,000 live births) and under-five mortality (94 deaths/1,000 live births). Nationally, roughly 5% of children under-five suffered from MAM and 2% from SAM, and 40% were stunted.[40] In Pujehun District specifically, rates of acute malnutrition were deemed to be above international thresholds for supporting SFP based on available data (facility data). However, over the course of the study, the team found rates of MAM were lower than previously reported (see **Section 6** for more information on sample size).

5. Formative Research

In July 2016, the FAQR team collaborated with Sierra Leone's DFN-MoHS and Ministry of Education, Science, and Technology (MEST), with support from the Sierra Leone offices of WFP and PPB, to conduct formative research prior to the start of the Four Foods Study. This activity explored locally standard or "proper" cooking methods in order to inform development of locally appropriate FBF cooking instructions for caregivers. This activity also helped the study team understand taste acceptability for CSB+ w/oil, CSWB w/oil, and SC+A. The details of that research are reported elsewhere.[41] In total, 96 female caregivers participated in standardized sensory and taste tests of one of

the three foods (32/food). Trained facilitators and home economics teachers engaged in controlled cooking observations of potential caregivers to gauge their ability to adhere to instructions. Focus group discussions provided additional insight into common porridge cooking practices.
The research team identified a standard and widely available plastic "baby feeding cup" for caregivers to use when measuring the flour. The baby feeding cup had an indented ring at a standard location on the cup. A single serving of SC+A was 136 g of flour, which equated to one full "baby feeding cup." A single serving of either CSB+ w/oil or CSWB w/oil required 86 g of flour, the equivalent to "flour filled just to the line" of the baby feeding cup. (See **Section 6** for more information on ration sizes.)

FVO was to be added to the CSB+ and CSWB flours at 26 g FVO for every 86 g of FBF. Three standard tablespoons of FVO proved accurate to arrive at this quantity. However, this measurement was identified before the plastic reusable oil bottles were procured. Once the oil bottles arrived, the team learned that "four bottle caps" of FVO was a more accurate measurement than three tablespoons. The local clinic team was thus trained to counsel caregivers to use "flour filled just to the line" and "four bottle caps of oil" for the CSB+ w/oil and CSWB w/oil study arms.

Porridge consistency was important to consumer acceptability and recipe adherence. As noted above, the addition of amylase to SC+ reduced the porridge's viscosity. Additionally, the longer the porridge cooked, the thinner it became (in contrast to the other FBFs)., although the thinning effect ceased after five minutes of cooking time Among Sierra Leone caregivers, thicker porridge is more culturally normal. Caregivers therefore required additional counseling to understand the counterintuitive nature of SC+A. If caregivers wanted a thicker porridge, study team members counseled them to reduce the quantity of water and not to increase the cooking time or the quantity of flour.

5.1 SBCC Program

From July to August 2017, the research team engaged in activities to develop counseling cards and training for PPB staff, community health workers (CHWs), and community-based lead mothers (LMs). The counseling cards were meant to improve understanding of how the foods should be prepared, consumed, and used within the context of healthy complementary practices. The operational theory behind the program was that by training health workforce leaders (i.e., CHWs, LMs, and PPB staff), those individuals would serve as advocates who would then reinforce positive behaviors among beneficiary caregivers while in the community, at home, or at the SFP.

The purpose of the SBCC formative research process was to understand community-level barriers and facilitators to recommended complementary food preparation practices in Pujehun District. The process also allowed the team to determine appropriate message delivery strategies to encourage change in feeding practices, when necessary. Ultimately, the research team included community and social mobilization methods such as weekly radio broadcasts, community talks, and community screening activities, as well as counseling cards tailored to the Pujehun context for use by community health workers, nurses, and lead mothers. These delivery strategies targeted factors such as knowledge, self-efficacy, and motivation, which were identified as either barriers or facilitators to effective use of the foods for MAM treatment.

The results of eight focus group discussions identified household members (e.g., mothers-in-law or husbands) who could be advocates for recommended preparation behaviors. The results also identified individuals (e.g., siblings under-5) who created challenges to sustaining recommended preparation and consumption patterns. From these discussions, counseling cards and radio messages were developed around four themes: ration size, ingredient quantities, dietary diversity in supplementary feeding, and strategies to reduce sharing. Once messages and instruction materials were created, training modules

were developed for local trainers, who then trained CHWs, PPB Clinic staff, and LMs. Further details of the formative research and focus group results can be found at foodaidquality.org.

6. Study Design

This study was a cluster-randomized, clinical effectiveness trial. The PHU sites where PPB screened, enrolled, and treated malnourished children were the clusters that served as units of randomization, each to one of the four SNFs. The CSB+ w/oil arm served as the comparison group because it was the SNF for MAM commonly programmed by FFP programs. **Figure 3** shows the locations of the selected PHUs in the study area and the arm to which each was assigned.

At each SFP visit, participants were given a two-week ration of the food randomized to that PHU and were told to return two weeks later to be evaluated. At each two-week visit, PPB staff took beneficiaries' anthropometric measurements, which were recorded on a paper clinic card, and provided another two-week food ration as illustrated in **Figure 1**. This continued until an outcome was reached for up to 12 weeks of rations distributed (whichever occurred first). Outcomes were defined as:

- Graduated (achieving MUAC ≥12.5 cm)
- Developed SAM (MUAC <11.5 cm or presence of bipedal edema)
- Death
- Default (missed three consecutive clinic visits)
- Failure (did not achieve MUAC ≥12.5 cm after receiving 12 weeks of rations)

The study also included mixed-methods data collected on subsamples of beneficiaries: structured interviews, qualitative focus group discussions, and direct observations both in beneficiaries' homes and at the PHU. (See **Section 6.4** for more detail)

Consent was obtained from beneficiaries' caregivers prior to their participation in the study; additional consent was obtained from the subsamples of caregivers selected for structured interviews, focus group discussions, or in-home observations. This study received ethical approval from the Tufts University Health Sciences Institutional Review Board, the WashU School of Medicine Institutional Review Board, and the Sierra Leone Ethics and Scientific Review Committee. The trial is registered at ClinicalTrials.gov: NCT03146897.

6.1 Sample Size and Sampling

For a cluster-randomized design comparing two proportions (i.e., one treatment arm to the comparison arm, CSB+ w/oil), planned sample sizes of 1,330 per arm (total of 5,320) sampled from seven sites per arm was calculated to achieve 80% power to detect a difference between the group proportions of 0.07 (7 percentage points). The test statistic used for this calculation was the two-sided Z-test (H0: P1 - P2 = 0; H1: P1 - P2 = D1 ≠ 0). The intracluster correlation used was 0.006 (based on prior data from Sierra Leone), and the significance level of the test was 0.05. This calculation assumed a baseline recovery rate of 70%. **Figure 4** shows the sample selection scenarios, assuming a graduation rate in the comparison group of 70%.

For this study, the sample size calculation was based on the primary outcome of "percent of children graduated from SFP." PASS 14 software was used for the calculations.[a]
However, for feasibility reasons, the total study sample size was capped at approximately 5,000 (i.e., 1,250 per arm). Increasing the number of PHUs (i.e., clusters) per arm allowed for a smaller detectable effect size with a lower total sample size (smaller n per site). Based on these scenarios, seven sites per

[a] PASS 14 Power Analysis and Sample Size Software (2015). NCSS, LLC. Kaysville, UT, ncss.com/software/pass.

arm (28 sites total) were chosen (see **Section 6.2**) as the most feasible option. An eighth site was later added to the CSWB w/oil arm due to low enrollment (described further below).

Section 4 (above) details why actual enrollment figures and graduation rates were lower than expected. During the study, the CSWB w/oil arm had lower overall enrollment numbers than the other three arms. To achieve better statistical power and balance among the arms, an eighth PHU site was added to the CSWB w/oil arm in January 2018. With sample sizes of between 579 and 768 per arm and graduation rates of between 62 and 65% per arm, the detectable effect size shifted from ~7 percentage points (as originally planned) to ~9 percentage points.

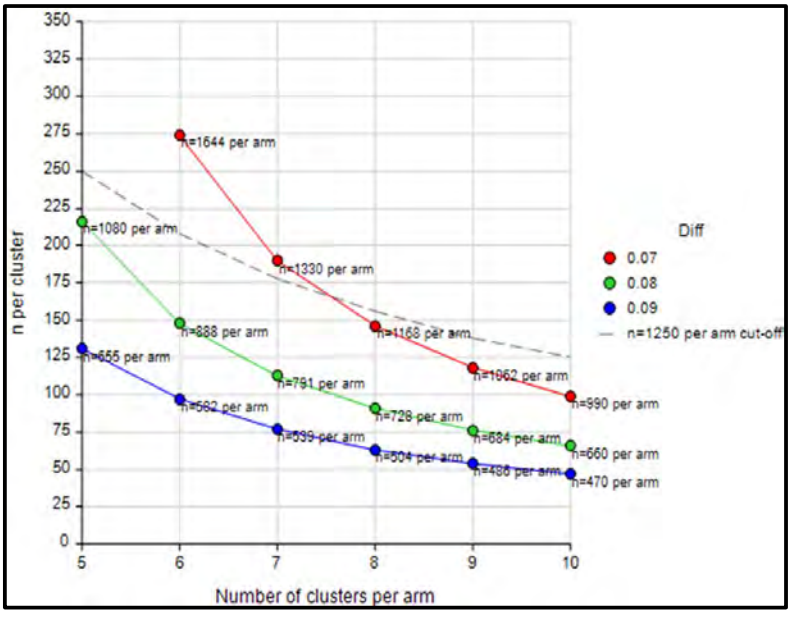
In addition to the primary analysis, three subsamples (not related to the biological sub-studies) of caregivers were selected to explore behavioral factors influencing effectiveness of the four foods. These sample sizes were chosen for logistical reasons: 420 in-depth interviews per study arm (maximum), of which 112 would include in-home observations (maximum). These subsamples were randomly chosen without replacement and were drawn from the list of beneficiaries who had received a study food within the last seven days and had not yet been interviewed.

A final subsample of caregivers was selected for participation in focus group discussions. To be selected for a focus group discussion, caregivers needed to have attended SFP within the last 90 days and have never been interviewed previously. Generally, the goal was to reach six caregivers for each focus group, but this was not always possible and so four was considered an allowable minimum.

6.2 PHU Selection and Randomization

There were 77 PHUs located throughout the peri-urban and rural chiefdoms of Pujehun District at the start of the study. A stratified block randomization technique was used to obtain four groups of seven PHUs; each group represented the variability of the district as equally as possible (**see Figure 3**). First, 60 PHUs with the highest 5-month accumulated MAM caseload were selected as the sampling frame based on data provided by the DHMT from the previous 12-month period. This list of 60 PHUs was then categorized according to the stratification criteria noted below, which were based on input from

Figure 4: Cluster and Total Sample Size Scenarios



community and government stakeholders. In the case that a selected PHU could not be included in the study (e.g., due to proximity to another study PHU or other logistical reasons), another PHU was selected from those remaining on the list.

PHU stratification criteria:

- MAM caseload
- Located within a town
- Chiefdom (all 12 chiefdoms of Pujehun District must be represented and fairly evenly distributed among the four arms)
- Type of PHU, i.e., Maternal and Child Health Post, Community Health Centers, and Community Health Posts (all three types of PHUs) must be represented and fairly evenly distributed among the four study arms
- Predominant mining PHU catchment areas (because livelihood can affect consumption patterns)
- Located on a major water body (indicates fisherman occupation)
- Presence of a current outpatient therapeutic treatment program (OTP) evenly distributed between the arms
- Accessibility (each arm containing some PHUs located either in Mano Sakrim chiefdom or on the upper/west side of the Moa river)

Details about the sampling process are provided in **Appendix 2.** The four groups were designated A, B, C, and D by Tufts. Then a researcher at WashU who was blinded to the PHU selection scheme randomly assigned the four groups into treatment arms by choosing a random order of four numbers (1,2,3,4) which were then decoded by Tufts.

6.3 Subjects' Eligibility and Enrollment

Eligible subjects were children 6-59 months old with MAM, defined by mid-upper arm circumference (MUAC) ≥ 11.5 cm and < 12.5 cm, without bipedal oedema, and having not received food rations from another organization for treatment of MAM. All eligible children who arrived at the PHU for evaluation and whose caregivers gave informed consent were included without regard to presence of chronic illness or permanent residence in the local community. Enrollment occurred on a rolling basis from April 2017 until September 2018. While participants were enrolled passively (i.e., based on whether they came to clinic), outreach was conducted in the surrounding communities by PPB staff and CHWs who periodically screened for MAM cases and made referrals to the PHUs. Children with MAM who were screened at a PHU that was not a study site were referred to the nearest study PHU for enrolment in the program. Screenings at non-study PHUs was conducted ad hoc (i.e., only when children arrived at the PHU and were actively screened; PHUs did not engage in community-based screening for MAM). In this way, although the program was resident in only 29 PHUs, the entire district was captured. If a child was found to be enrolled twice or at two separate PHUs, the child was removed from the study but still actively treated for MAM as part of the SFP.

6.4 Data Collection Instruments

This mixed method study collected quantitative and qualitative data. Specific data collection instruments are summarized in **Table 2.** KoBoToolbox[b] was used for electronic data collection and overall data management.

[b] https://www.kobotoolbox.org/

The SFP-specific data, which included anthropometric and morbidity information as well as the demographic enrollment questionnaire, were collected by PPB staff on paper clinic cards and entered in KoBoToolbox webforms by trained Caritas Bo data clerks. Every clinic card was entered twice (once by each of the two data clerks) within 48 hours of the caregiver arriving at clinic. Every month, the total database of newly enrolled beneficiaries was downloaded from the KoBoToolbox cloud and compared against the logged informed consent forms. Any discrepancies were resolved during the first week of every month. (See **Figure 1** for more detail.)

The field surveys and observational data collection were conducted by the team at Caritas Bo and were collected using Samsung Galaxy A6 tablets with the KoBoCollect data collection app. Each tablet was equipped with a SIM card in which mobile data were loaded, and all surveys were uploaded to the web-based cloud from the field. Field notes from the in-home observations were typed by a group of professional typists. Focus groups were recorded using either handheld recorders or tablets. Recordings were then translated and transcribed by hand by two different research assistants (RAs) working separately. The final transcripts were then compared against each other for discrepancies by a supervisor. Any discrepancies were resolved by reviewing the recording and two supervisors agreeing on the correct translation. The agreed upon translation was then typed by the typists.

Table 2: Data Collection Tools for Each Thematic Area of the Four Foods Study

Effectiveness	Cost-effectiveness	Influencing Factors
Bi-weekly anthropometry	Clinic Day observations	Key Informant Interviews
Bi-weekly morbidity and treatment notes	Warehouse observations	Individual interviews with recipients, lead mothers, and HNPs
Socio-economic surveys at enrollment	Individual interviews with CHWs and lead mothers	In-home observations with recipients
Community questionnaires	In-home observations with recipients	Focus group discussions with recipients and distribution committees
	Financial records	

7. Qualitative Methods

Transcripts were analyzed using a thematic analysis that was developed during the study's design and focused on behaviors broadly related to use of the study foods. The analysis was guided by questions in the interview guide as well as relevant literature.[42] QSR International's NVivo version 12 was used for data management and analysis. [c] Development of the codebook by two analysts began with line-by-line coding of three randomly selected focus group discussion (FGD) transcripts from the list of 25 completed FGDs. The codebook included parent codes, child codes, and definitions that materialized through consensus between the two coders. Any discrepancies were discussed and resolved by mutual agreement. This iterative process was used to develop the finalized codebook.

[c] NVivo qualitative data analysis Software. QSR International Pty Ltd., Version 12, 2019.

After resolving discrepant codes, one analyst led coding for the remaining focus group discussions. The finalized codebook was revised iteratively as new concepts arose throughout this process. When a new code was identified, all prior FGDs were recoded using the amended codebook. This coding process continued until similar concepts and topics were apparent in all FGDs, known as theoretical saturation.[43] Axial coding was used to connect themes relating to factors influencing the way the study foods were oded text was displayed in a matrix and stratified by study arm to connect themes among the study foods. To ensure coding consistency, intercoder reliability checks were conducted once by two researchers on the team. A Kappa score average of 99.6% was obtained by randomly selecting one FGD and having the two researchers code the same transcript using the final codebook.

8. Effectiveness

8.1 Methods

a. Effectiveness

The primary analysis aimed to determine whether differences in recovery rates existed among children treated with one of four SNFs. A conceptual framework was developed that included individual child level and household level characteristics. Using this framework, descriptive statistics were calculated and stratified by study arm. Variables identified as potential confounders or predictors using a stepwise approach were selected for inclusion in the regression models as covariates. Mixed-effects logistic regression models were fit using PHU as the random effect. The outcome was defined as achieving MUAC ≥ 12.5 cm within 12 weeks versus not (see **Section 6** for more information). Post-estimation marginal predictions for the study arm variable were obtained and used for the cost-effectiveness evaluation. In addition to the main outcome, a survival analysis was conducted to assess the time to graduation among the subsample of those who graduated from the MAM treatment SFP (n=1,676). Cox Proportional Hazards Models were built to obtain hazard ratios to represent the rate of graduation per unit time (days) for each study arm compared to CSB+ w/oil. A more detailed description of the modeling procedures is in **Appendix 2**.

Secondary outcomes included weight gain velocity (g/kg/day) and total MUAC gain (in cm) throughout the treatment period. These were tested for differences between arms using mixed-effects models adjusted for child's age, sex, baseline weight-for-length Z-score (weight gain model) or MUAC (MUAC gain model), and number of rations distributed.

b. Sustained recovery at one-month follow-up

A secondary analysis, exploratory in nature as a secondary research question, determined whether differences in sustained recovery exist among graduated children treated with one or another of the four SNFs. The same process used to structure the main effectiveness analysis was used to analyze the incidence of sustained recovery at one-month post-graduation. The one-month post-recovery time period was chosen because it was the time point with the most complete data and closest to outcome so as to capture food-specific differences. As shown in **Table 3**, at the three- and six-month follow-up appointments, the proportion of missed visits was deemed too high for reliable analysis. Descriptive statistics were calculated and stratified by study arm. Variables selected for inclusion in the final model as covariates were those considered potential confounders or predictors.

The binary outcome variable was defined as not having sustained recovery if one of two conditions were met: the MUAC at the one-month follow-up appointment was <12.5 cm or the nurse recorded "relapse" on the clinic card, indicating that a child had relapsed between graduation and the one-month follow-up appointment. Crude and adjusted mixed-effect logistic regression models were fit using PHU

as the random effect. Observations with missing covariate data were dropped from the models. To evaluate whether missed visits had an impact on the parameter estimates, three models were compared: 1) excluding those with missed visits at one month post-graduation, 2) categorizing the outcomes of those with missed visits at one month post-graduation as all relapsed, and 3) categorizing the outcomes of those with missed visits at one month post-graduation as all sustained recovery. The marginal predictions estimated from these models for each study arm were calculated and used for additional cost-effectiveness evaluation.

Data were analyzed using Stata 15 (StataCorp, College Station, TX).

Table 3: Data Collection Tools for Each Thematic Area of the Four Foods Study

	CSWB w/oil n (%)	SC+A n (%)	CSB+ w/oil n (%)	RUSF n (%)	Total n (%)
Sustained Recovery	267 (74%)	294 (71%)	317 (75%)	313 (66%)	1191 (71%)
Relapsed by 1-month	62 (19%)	79 (19%)	82 (19%)	116 (24%)	339 (20%)
Missed the 1-month Visit	31 (9%)	42 (10%)	25 (6%)	48 (10%)	146 (9%)
Missed the 3-month Visit	135 (38%)	155 (37%)	125 (29%)	206 (43%)	621 (37%)
Missed the 6-month Visit	231 (64%)	256 (62%)	233 (55%)	323 (68%)	1043 (62%)

8.2 Results

a. Effectiveness

A total of 2,691 children were consented and enrolled; 38 were excluded; and 2,653 were included in analysis. A flowchart in **Figure 5** shows enrollment and outcome figures by study arm. Reasons for exclusion included:

- Enrollment error, i.e., MUAC ≥ 12.5cm or < 11.5 cm at start (n=10)
- Known to be younger than 5.5 months or older than 59.5 months at enrollment (n=14)
- Child had Cerebral palsy (n=1)
- Error with assigned outcome or child progressed to SAM status because of a twin diagnosed with SAM (n=7)
- No recorded outcome or no outcome reached before the study ended (n=4)
- Incomplete information to complete enrollment in study (n=2)

Table 4 shows descriptive characteristics of the sample at enrollment by study arm. Overall, the study arms were balanced. Some differences were noted and controlled for in analysis, including:

- Sex
- Whether the child was transferred into the study from SAM (i.e., received prior treatment for SAM and transferred to SFP when MUAC increased to 11.5 or greater)
- Whether the child had a twin
- Caregiver's age

- Number of males age 0-5 years in the household
- Baseline weight-for-height Z-scores
- Baseline morbidities

Additional covariates that were related to the outcome in univariate models included:

- Child's age
- Caregiver's level of education
- Season of enrollment
- Presence of females age 65 or older
- Number of females age 0-5 years
- Socioeconomic index (radio, mattress with bed, motorcycle/scooter, agricultural land)
- Household toilet type
- Unprotected well as the main drinking water source
- Zinc roof type

There were few missing data overall and within each study arm, with <2% missing for each variable. Overall, 63.2% of treated children graduated from the SFP. By study arm, the graduation rates ranged from 62.1% (RUSF arm) to 64.5% (SC+A arm) see **Figure 5**. In all models, there was no significant difference in outcomes by study arm. The adjusted marginal predictions for graduation from MAM treatment within 12 weeks are displayed in **Figure 6**. Due to collinearity with the other anthropometric-related covariates, the average length and the average weight of the 3 measurements taken per child per visit as well as the weight-for-age Z-score at baseline were dropped from the adjusted model.

Table 4: Child and Household Characteristics at enrollment[1]

Child & Household Characteristics at Enrollment	CSWB (n=579)		SC+ (n=643)		CSB+ (n=663)		RUSF (n=768)		Total (n=2653)	
	n or mean±SD	*% or med(min,max)*	*n or mean±SD*	*% or med(min,max)*	*n or mean±SD*	*% or med(min,max)*	*n or mean±SD*	*% or med(min,max)*	*n or mean±SD*	*% or med(min,max)*
Female sex	329	57%	399	62%	365	55%	430	56%	1,523	58%
Age, in months	13.2±7.9	10.7 (5.8,59.5)	13.2±8.1	10.5 (5.8,58.1)	13.1±7.7	10.5 (5.7,55.6)	13.5±8	11.1 (5.6,59.3)	13.3±7.9	10.7 (5.6,59.5)
Breastfeeding	453	80%	510	80%	523	80%	595	79%	2,081	80%
Season of enrollment										
Rainy season	363	63%	421	66%	399	60%	441	57%	1,624	61%
Transferred from SAM	114	20%	174	27%	166	25%	218	28%	672	25%
Twin at birth	23	4%	32	5%	39	6%	20	3%	114	4%
Mother is caregiver	546	95%	589	92%	608	92%	700	91%	2,443	92%
Caregiver is married	490	85%	541	85%	571	86%	647	85%	2,249	85%
Caregiver's age, in years	28.1±8.2	27 (16,80)	27.1±7.4	25 (17,65)	28.3±8.4	27 (17,65)	27.5±8.2	26 (16,69)	27.7±8.1	26 (16,80)
Level of education										
None	321	56%	340	54%	369	56%	389	51%	1,419	54%
Some or completed primary	141	24%	146	23%	146	22%	200	26%	633	24%
Some, completed, or more than secondary	116	20%	150	24%	145	22%	176	23%	587	22%
SES and HFIAS										
HFIAS category (mild/moderate collapsed)										
Food Secure	134	23%	148	23%	178	27%	220	29%	680	26%
Mildly or Moderately Food Insecure Access	80	14%	93	15%	77	12%	124	16%	374	14%
Severely Food Insecure Access	365	63%	395	62%	406	61%	424	55%	1,590	60%
SES Wealth Index (quintiles of score derived from PCA)										
Lowest	123	22%	115	18%	122	19%	162	21%	522	20%
Mid-Low	138	24%	105	17%	115	18%	165	22%	523	20%
Medium	110	19%	122	20%	123	19%	164	22%	519	20%
Mid-High	111	19%	133	21%	145	22%	134	18%	523	20%
Highest	89	16%	152	24%	148	23%	135	18%	524	20%

[1] Percentages are of non-missing values, < 2% missing for each variable

Figure 5: Reasons for Exclusion and Enrollment and Outcome Figures by Study Arm

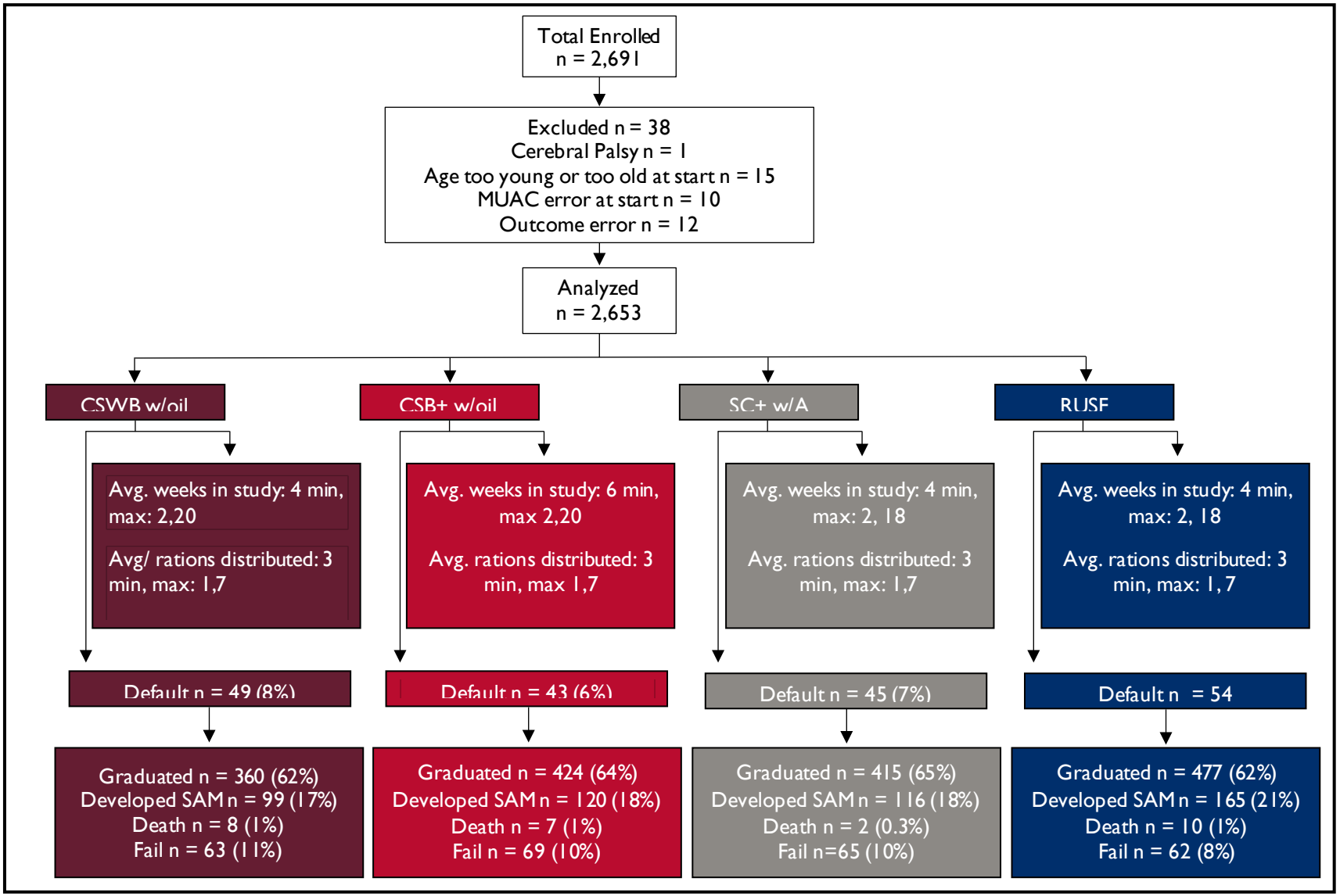
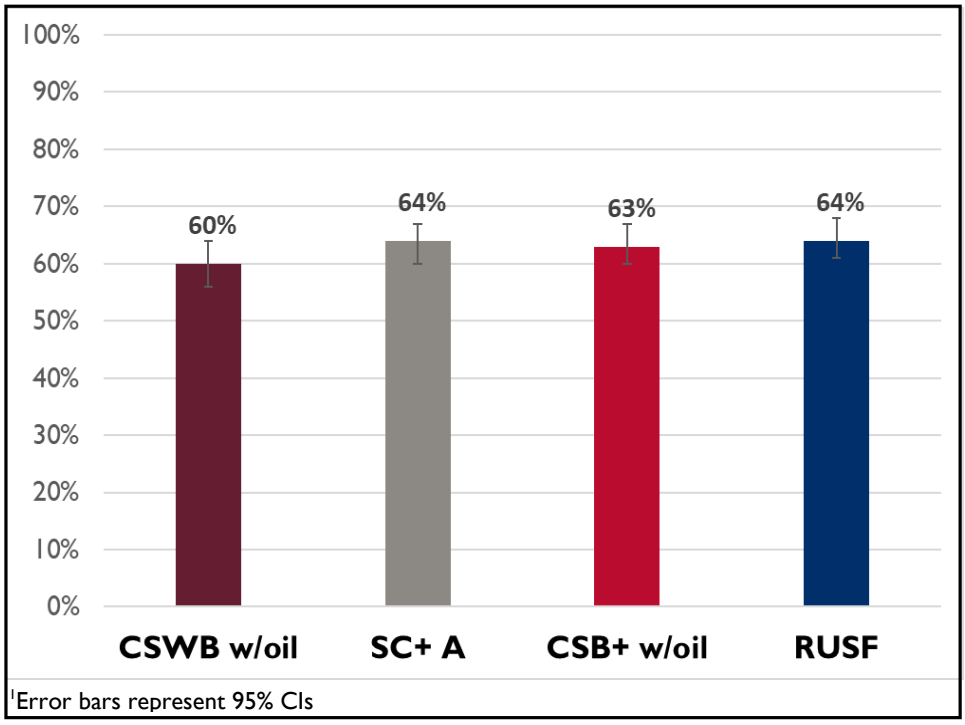


Figure 6: Percentage Graduated Within 12 Weeks, Adjusted Marginal Predictions[1]



[1]Error bars represent 95% CIs

Several variables were significantly associated with graduation from SFP. Children with the following enrollment characteristics had significantly higher odds of graduation: higher average MUAC, higher length-for-age Z-score, enrolled in the rainy season, presence of females age 65 or older within the household, and presence of diarrhea within prior two weeks. Among those who were transferred from SAM, having a zinc roof type and having had a fever within the prior two weeks were associated with significantly lower odds of graduation. Odds ratios and 95% confidence intervals from the adjusted model are displayed in **Table 5.**

Mean time to graduation (for those that graduated) was 35.8 days overall (SD=22.9) and ranged from 34.9 to 36.6 days by study arm, with the shortest mean time to graduation in the RUSF arm and the longest in the CSB+ w/oil arm. However, none of these differences was significant. **Table 6** shows the Hazards Ratios (HR) from Cox Proportional Hazards models for time to graduation, and **Figure 7** shows Kaplan-Meier curves for time to graduation by study arm.

Average weight gain velocity (g/kg/day) throughout the study period ranged from 12.6 in the CSB+ w/oil arm to 14.1 in RUSF (12.8 and 13 in CSWB w/oil and SC+A, respectively), with no statistically significant differences among arms. Similarly, average MUAC gain throughout the study period was the same for all four study arms, at 0.4 cm with no statistically significant differences detected.

Table 5: Odds Of Graduation From SFP[1], Adjusted Mixed-Effects Logistic Regression

	OR	95% CI	P-value
CSB+	Ref.		
CSWB w/oil	0.83	(0.64, 1.08)	0.16
SC+ w/A	1.01	(0.78, 1.3)	0.95
RUSF	1.05	(0.82, 1.34)	0.69
Child's sex (male)	1.26	(0.97, 1.64)	0.09
Child's age (in months)	1.02	(1, 1.04)	0.07
Breastfeeding	0.93	(0.69, 1.26)	0.65
Enrolled during dry/hot season	Ref.		
Rainy season	1.55	(1.29, 1.87)	<0.01
Child transferred from SAM	0.56	(0.45, 0.7)	<0.01
Received supplementary food last month	0.90	(0.62, 1.31)	0.59
Child had twin at birth	1.11	(0.7, 1.74)	0.66
Caregiver's age	1.01	(0.99, 1.02)	0.21
No formal education	Ref.		
Some or completed primary	0.93	(0.74, 1.16)	0.51
Some, completed, or more than secondar	0.98	(0.76, 1.27)	0.91
Child has no living siblings	Ref.		
One	0.89	(0.69, 1.16)	0.40
Two	1.20	(0.9, 1.6)	0.20
Three or more	1.04	(0.78, 1.38)	0.79
No males aged 0-5 years in household	Ref.		
One	0.94	(0.74, 1.19)	0.59
Two or more	0.97	(0.74, 1.28)	0.83
No females aged 0-5 years in household	Ref.		
One	0.79	(0.61, 1.04)	0.09
Two or more	0.82	(0.61, 1.1)	0.18
One or more females aged 65+ in household	1.43	(1.08, 1.9)	0.01
Radio	1.04	(0.87, 1.26)	0.65
Mattress with bed	1.00	(0.82, 1.23)	0.98
Motorcycle/scooter/okada	0.97	(0.74, 1.26)	0.81
Agricultural land	0.94	(0.74, 1.2)	0.63
Private household toilet	Ref.		
Shared/public	0.96	(0.79, 1.17)	0.71
Unprotected well as drinking water source	1.52	(0.94, 2.48)	0.09
Zinc roof type	0.81	(0.66, 1)	0.05
Average MUAC at enrollment	11.76	(8, 17.29)	<0.01
Length-for-age Z-score at enrollment	1.17	(1.06, 1.28)	<0.01
BMI-for-age Z-score at enrollment	1.27	(0.85, 1.9)	0.24
Weight-for-length Z-score at enrollment	0.99	(0.68, 1.44)	0.96
Fever at enrollment (within prior 2 weeks)	0.78	(0.61, 0.99)	0.04
Diarrhea at enrollment (within prior 2 weeks	1.67	(1.12, 2.48)	0.01
Cough at enrollment (within prior 2 weeks)	1.15	(0.89, 1.49)	0.27

[1]Graduation defined as reaching a mid-upper-arm circumference ≥ 12.5 cm within 12 weeks

Table 6: Cox Proportional Hazards Models Assessing Time to Graduation by Study Arm

Graduation rate per day[1]	Crude HR	95% CI	Adjusted[2] HR	95% CI
	n=1,676		*n*=1,644	
Study arm (Ref = CSB+ w/oil)				
CSWB w/oil	1.03	0.90, 1.19	1.02	0.89, 1.18
SC+ A	1.02	0.89, 1.17	1.04	0.91, 1.20
RUSF	1.05	0.92, 1.19	1.09	0.96, 1.25

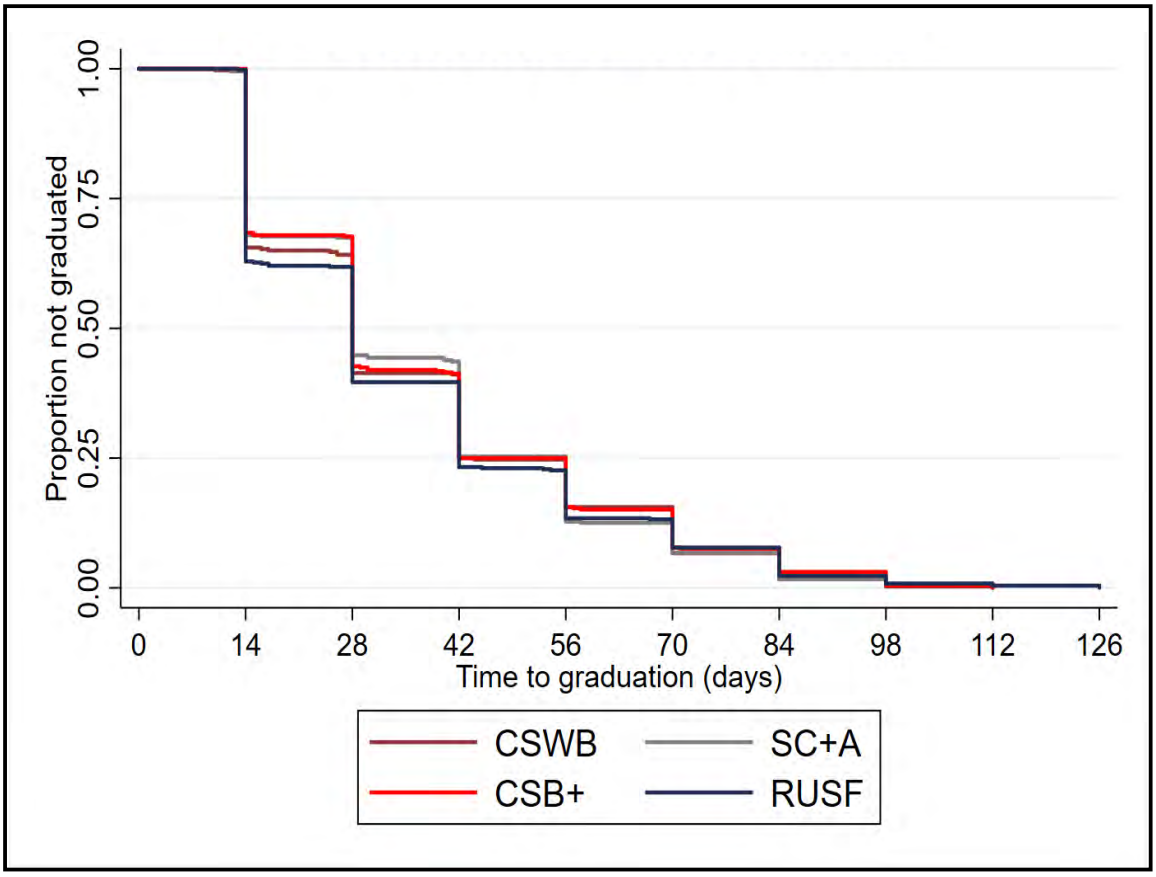
Notes: CSB+ = Corn Soy Blend Plus, CSWB = Corn Soy Whey Blend, SC+ A = SuperCereal Plus with amylase, RUSF = Ready-to-Use Supplementary Food

[1]Graduation defined as reaching a mid-upper-arm circumference ≥ 12.5 cm within 12 weeks

[2]Adjusted models control for: gender, breastfeeding, season, transfer from program for severe-acute-malnutrition, child age, weight at enrollment, muac at enrollment, bmi z-score at enrollment

*p<0.05

Figure 7: Kaplan-Meier Survival Estimates: Time to Graduation by Study Arm



b. Sustained recovery at one-month follow-up

A total of 1,676 beneficiaries graduated from the SFP based on MUAC. Of these, most (91.3%) returned to clinic for a one-month follow-up appointment. The number of recovered beneficiaries that missed the one-month follow-up visit was proportionately lower among CSB+ w/oil beneficiaries (5.9%) than among the other three arms. **Table 3** illustrates differences in attendance at one-month post-graduation by arm.

In unadjusted analysis, all 1,531 beneficiaries that came for a one-month follow-up appointment were included; individuals that missed the one-month appointment were excluded. As covariates were included, beneficiaries with incomplete data were dropped, for a final sample of 1,489 in adjusted analysis. In total, 22.14% of the 1,531 beneficiaries that came for one-month follow-up relapsed to MAM. In both adjusted and unadjusted models, there were no statistically significant differences in the odds of sustained recovery between any of the FBFs and CSB+ w/oil. A significant difference between RUSF and CSB+ w/oil existed in both adjusted and unadjusted models, though this was attenuated in the adjusted model. Overall, the odds of sustained recovery were lower for RUSF than for CSB+ w/oil. **Figure 8** displays the adjusted marginal predictions for both adjusted and unadjusted models.

Incidence of fever, vomiting, diarrhea, and cough at graduation were found to be collinear; only incidence of diarrhea was therefore chosen for inclusion as a proxy for comorbidities at graduation. Variables associated with significantly higher odds of sustained recovery in the adjusted model included the child's age, weight gain velocity over treatment period, MUAC at graduation, graduation in rainy season, and being severely food insecure at enrollment.

Figure 8: Percentage Sustained Recovery at 1-month Follow-up, Adjusted Marginal Predictions[1]

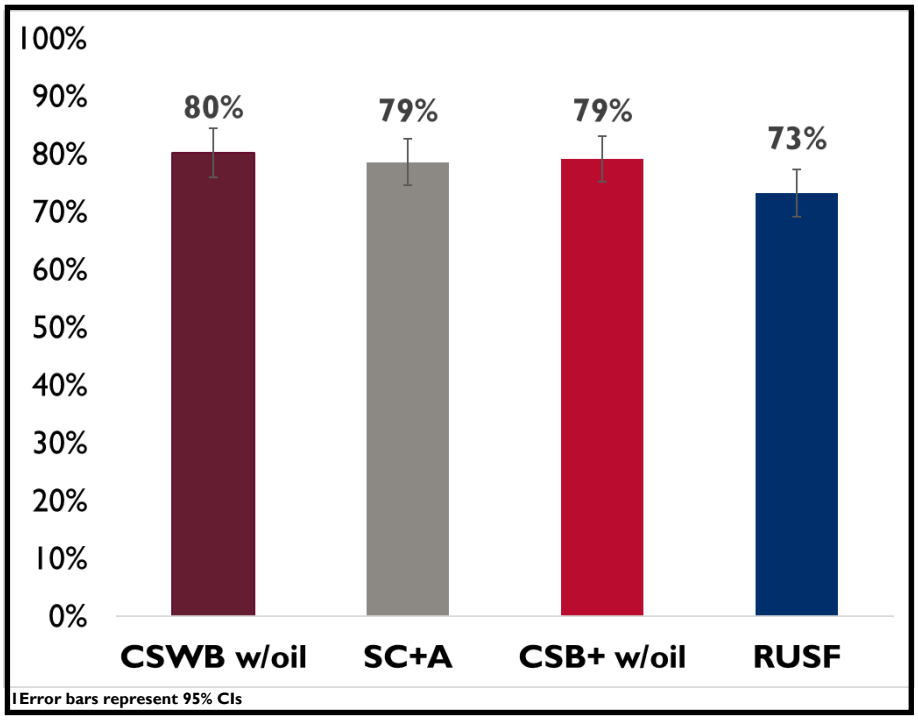


Table 8: Odds of Sustaining Recovery[1] After Graduation from SFP, Adjusted Mixed Effects Logistic Regression

	Unadjusted			Adjusted (exclude missed)			Missed visits = relapse			Missed visits = sustained		
	OR	**95% CI**	**P-value**	**OR**	**95% CI**	**P-value**	**OR**	**95% CI**	**P-value**	**OR**	**95% CI**	**P-value**
CSB+ w/oil (ref.)												
CSWB w/oil	1.11	(0.77, 1.61)	0.57	1.08	(0.73, 1.6)	0.71	0.94	(0.67, 1.34)	0.75	1.12	(0.76, 1.65)	0.56
SC+A	0.96	(0.68, 1.36)	0.83	0.96	(0.67, 1.4)	0.85	0.85	(0.61, 1.19)	0.35	0.98	(0.68, 1.41)	0.91
RUSF	0.70	(0.51, 0.96)	0.03	0.70	(0.49, 0.99)	0.04	0.64	(0.46, 0.88)	0.01	0.74	(0.52, 1.04)	0.08
Child's age (in months)				1.04	(1.02, 1.06)	0.00	1.02	(1, 1.03)	0.01	1.04	(1.02, 1.06)	0.00
Breastfeeding				0.995	(0.98, 1.01)	0.55	0.993	(0.98, 1.01)	0.39	0.996	(0.98, 1.01)	0.62
Child transferred from SAM				0.996	(0.97, 1.02)	0.71	0.993	(0.97, 1.01)	0.51	0.997	(0.97, 1.02)	0.77
Child admitted to hospital in 2 weeks preceding enrollment				1.57	(0.71, 3.46)	0.26	1.55	(0.8, 3)	0.19	1.48	(0.68, 3.24)	0.32
Diarrhea at graduation (within prior 2 weeks)				0.94	(0.65, 1.37)	0.75	1.01	(0.99, 1.02)	0.47	0.94	(0.64, 1.37)	0.73
Weight gain velocity during treatment period (kg/day)				1.03	(1.01, 1.04)	0.00	1.02	(1.01, 1.03)	0.00	1.02	(1.01, 1.04)	0.00
Average MUAC at graduation				15.03	(6.5, 34.75)	0.00	3.93	(2.14, 7.21)	0.00	15.59	(6.82, 35.63)	0.00
Caregiver's Age				1.002	(0.99, 1.01)	0.55	1.004	(1, 1.01)	0.26	1.002	(0.99, 1.01)	0.59
Household Food Insecurity												
Moderately Food Insecure				1.18	(0.78, 1.8)	0.43	1.15	(0.8, 1.65)	0.46	1.20	(0.79, 1.8)	0.39
Severely Food Insecure				1.38	(1.02, 1.85)	0.03	1.28	(0.99, 1.67)	0.06	1.40	(1.05, 1.87)	0.02
Socioeconomic Status												
Mid-Low				0.73	(0.49, 1.1)	0.13	0.86	(0.61, 1.23)	0.42	0.73	(0.49, 1.09)	0.12
Medium				1.01	(0.67, 1.52)	0.97	1.13	(0.79, 1.63)	0.50	0.96	(0.64, 1.44)	0.85
Mid-High				1.09	(0.71, 1.66)	0.70	1.01	(0.71, 1.46)	0.94	1.08	(0.71, 1.64)	0.73
Highest				0.89	(0.58, 1.37)	0.61	0.80	(0.56, 1.15)	0.24	0.95	(0.63, 1.44)	0.80
Season of Graduation												
Rainy season				1.39	(1.05, 1.82)	0.02	1.29	(1.01, 1.63)	0.04	1.35	(1.03, 1.77)	0.03

1 Sustained Recovery defined as maintaining a MUAC ≥12.5 for 1-month after graduation

As detailed in **Table 3**, the proportion of missed visits at the one-month follow-up visit varied across arms. To understand whether these differences in attendance at the one-month follow-up could have influenced conclusions of the analysis, sensitivity analysis was conducted. **Table 8** details adjusted models for two scenarios: [A] all missed visits categorized as having relapsed to MAM or [B] all missed visits categorized as having sustained recovery post-graduation. The direction of the effect in both models was consistent with the main model that excluded missed visits. While RUSF had significantly lower sustained recovery than CSB+ w/oil in the main model and in model B, it was only marginally significant (p=0.08) for model A.

8.3 Summary

The four SNFs performed similarly across measures of effectiveness. Though the total sample size was lower than expected, the effectiveness differences between the foods were small, making it unlikely that a statistical difference would have been detected even if the full sample size had been achieved. Compared to CSB+ w/oil, the odds of sustained recovery at one-month post-treatment were lowest among children receiving RUSF.

9. Costs & Cost-effectiveness

9.1 Methods

a. Costs

Five costing perspectives are shown in **Figure 9**, with more detail found in **Appendix 3**. The figure details costs incurred by different stakeholders related to implementation of the SFP. Two combined perspectives were used in the cost-effectiveness analyses: the *program perspective* (combining donors, implementers, government, and lead mother volunteers [i.e., entities associated with the provision of services]) and the *combined program and caregiver perspective,* which incorporates the cost to caregivers of participating in the program.

Figure 9: Five costing perspectives and the included stakeholders

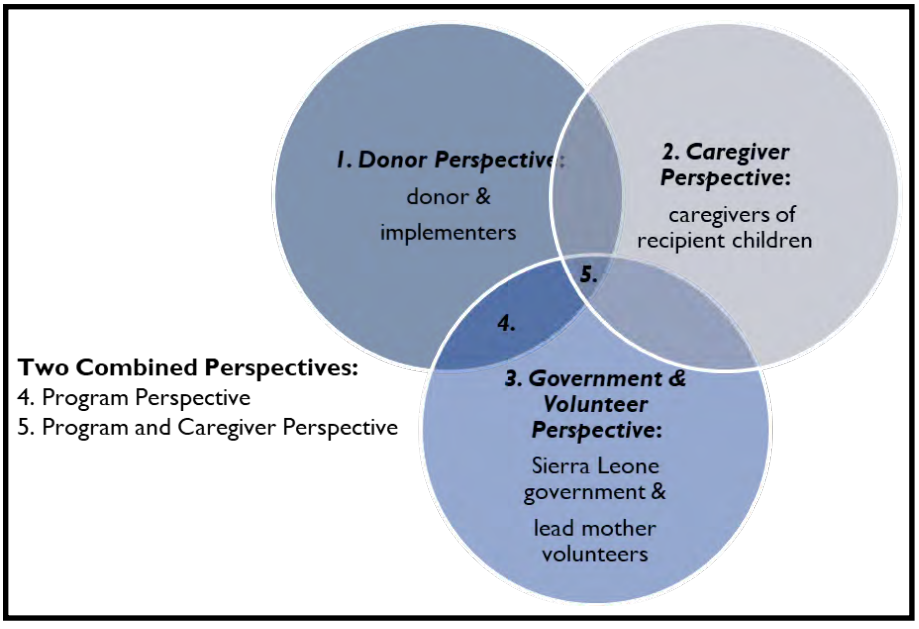


Figure 10: Cost components, summary cost measures, and cost-effectiveness measures

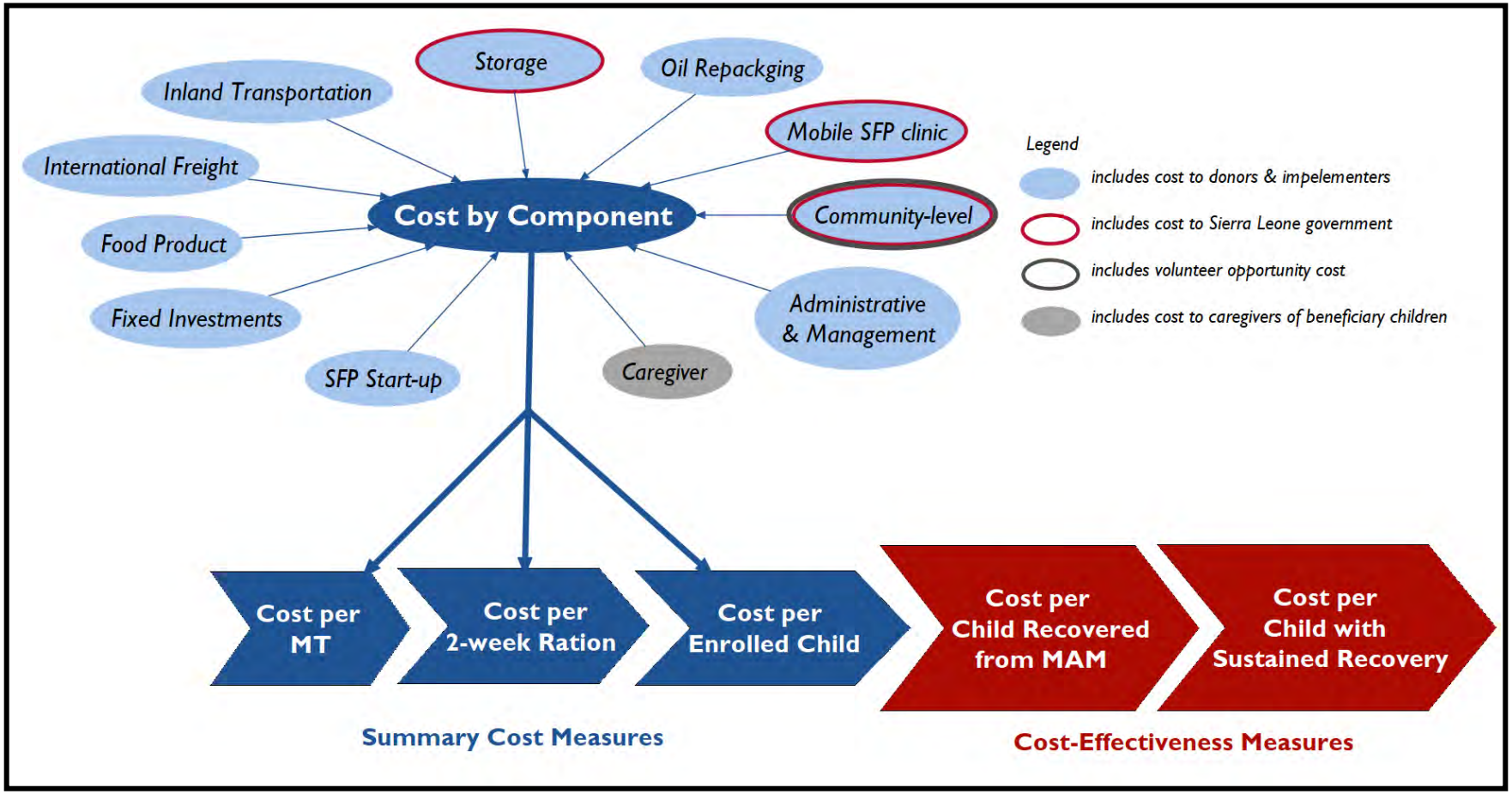


Table 9: Definition of cost components and corresponding data sources

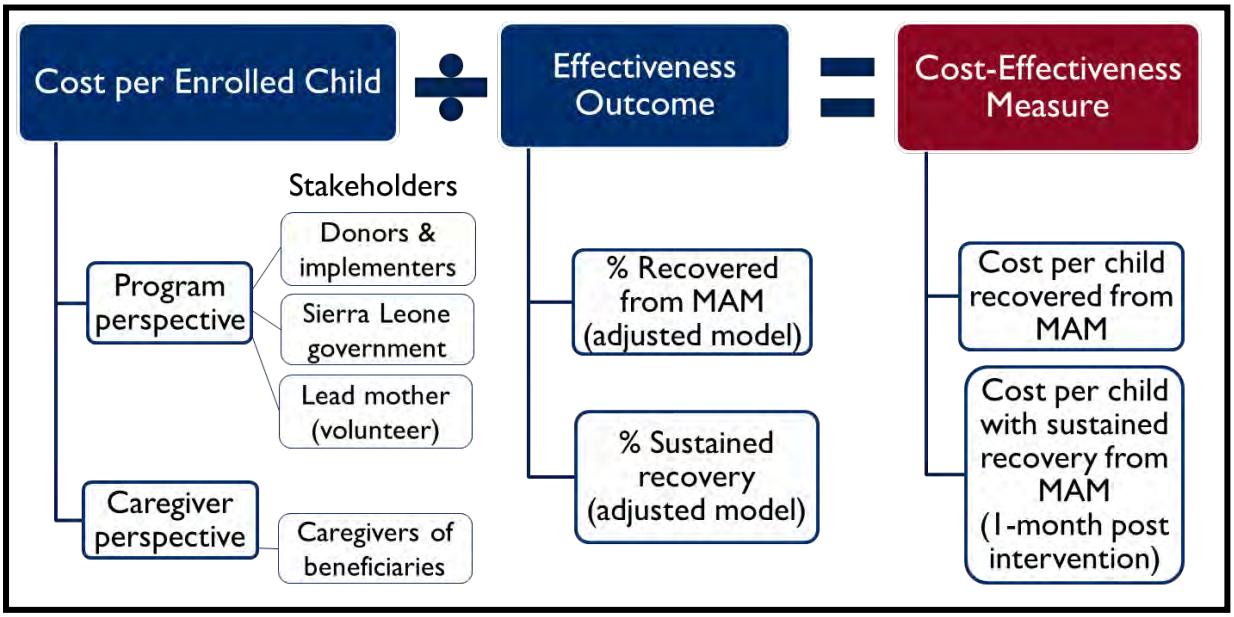
Cost Component	Definition	Study Instruments & Data Sources
Start-up Activities	Cost to set up the supplementary feeding program during the start-up period (between 09/2016 and 03/2017), including setting up Pujehun food storage, clinic scoping, supplies purchasing, clinic staff training, community health worker outreach.	Accounting records from PPB- Sierra Leone
Fixed Investment	Cost of fixed investments for the supplementary feeding program, including equipment/technology, fixed materials, vehicles, and PPB Pujehun office space.	Accounting records from PPB- Sierra Leone
Food Product	Cost of the specific specialized nutritious food and fortified vegetable oil	Realistic price quotes for experimental products from Didion Milling; Historical procurement price data for standard products from FFP and WFP
International Freight	Cost of international shipping from the United States to Freetown, Sierra Leone	Realistic quotes from BKA Logistics
Inland Transportation	Cost of food transportation and loading/unloading from Freetown to Kenema WFP Warehouse, and from Kenema Warehouse to Pujehun Storage. (transportation and labor)	Observation[1] (2) instruments from the study; Accounting records from WFP-Sierra Leone and PPB- Sierra Leone
Storage	Cost of storing the foods at the WFP warehouse in Kenema and at the Pujehun Storage. (government-provided space, labor, relevant services, and variable supplies)	Accounting records from WFP-Sierra Leone and PPB- Sierra Leone
Repackaging (fortified vegetable oil ONLY)	Cost of repackaging fortified vegetable oil 4L cans into small bottles (labor and materials)	Observation[1] (2) instruments from the study; Accounting records from PPB- Sierra Leone
Mobile Supplementary Feeding Program Clinic Activities	Cost of staff and food transportation between PPB Pujehun office and 28 PHUs and distribution of food and other MAM treatment activities at the PHUs. (labor, accommodation, transportation, and variable supplies)	Observation[1] (~2 per PHU for 28 PHUs) instruments from the study; Accounting records from PPB- Sierra Leone, and government salary information
Community-level Activities	Cost of conducting community-level activities including IEC, screening of MAM cases, outreach for missed visits, training etc. (labor cost of staff and CHWs, and opportunity cost of lead mothers' time)	Interview (193 CHWs and 178 lead mothers) instruments from the study; Accounting records from PPB- Sierra Leone, and government salary information
Administrative and Management Activities	Cost of managing the supplementary feeding program (labor, transportation, relevant services, and variable supplies)	Accounting records from PPB- Sierra Leone
Caregiver Activities	Monetary and opportunity cost of caregivers' time participating in the program, including transportation spending and time, clinic time, time for community activities, and household study food preparation/serving and feeding time	Observation (373 caregivers at PHUs and 269 caregivers at households) and interview (957 caregivers, 193 CHWs, and 178 lead mothers) instruments from the study

CHW: community health workers; MAM: Moderate Acute Malnutrition; IEC: Information, Education, and Communication; L: liter; PHU: Peripheral Health Unit; PPB: Project Peanut Butter; WFP: World Food Programme

[1] *In most cases, one observation took place in 2017 and the other took place in 2018.*

An activity-based with ingredients costing approach was used which identified 11 cost components.[44] These cost components are shown in **Figure 10** and defined in **Table 9. Figure 10** describes in more detail which stakeholders contributed to each cost component. Cost data were collected using a variety of study instruments and data sources (see **Table 9**). In addition to using data on costs actually incurred in the study, more realistic cost estimates for product procurement and international freight were obtained externally from historical cost data or price quotes from suppliers so that data could be generalized to other contexts. This was necessary because two of the foods were novel and procured only for the present study; the unit price was higher for the small quantities required than would have been the case if the foods had been ordered for a full-scale program.
Details on how product and international freight costs were derived are explained in the footnotes of **Table 10** and **Table 11**.

Figure 11: Methods Used to Assess Cost-Effectiveness for Each Arm



Cost data for each component were summarized into one of three summary cost measures for each arm (**Figure 10**):

- Cost per MT
- Cost per biweekly ration
- Cost per enrolled child

For each arm, cost per MT and cost per biweekly ration were converted to cost per enrolled child by either biweekly ration size or number of rations received. Differences in adherence to food collection or treatment duration by arm were accounted for in the cost comparisons.

Total cost per enrolled child was calculated by summing the cost components for each of the five costing perspectives separately. To represent variation in total cost from the caregiver perspective, uncertainty ranges were constructed using distributions of time for activities that differed by arm (e.g., preparation time or time spent at the first clinic visit). (See **Section 7.2** and **Appendix 3** for more details.)

All cost results presented in "$" are 2018 USD based on annual exchange rates between Sierra Leone leones (SLL) and USD and adjusted for inflation (annual rates of US GDP implicit deflator). The opportunity cost for lead mothers and caregivers was valued at $0.38 per hour based on the Sierra Leone minimum wage of SLL 500,000 per month, assuming five workdays per week, eight hours per day, and 10 national holidays per year.

Each of the three FBFs procured for this study was in some way a novel specification (in packaging and/or formulation), while RUSF and FVO were standard specifications. The FBFs and FVO were imported from the United States, while RUSF was procured locally in Sierra Leone. Novel formulations could affect the product cost, while novel packaging and procurement channels could affect both product cost and international freight cost. To explore the cost implications of different specifications and procurement channels in comparable conditions across arms, four policy-relevant scenarios were constructed and described in **Appendix 3**. Scenario 1 served as the primary scenario used for cost-effectiveness analysis for the following reasons:

- Package specifications other than those used in the study could result in an unobserved impact on effectiveness.
- Assuming the same procurement channel improves comparability across arms.
- Data sources for imported products are more available and more comparable than data sources for Local and Regional Procurement (LRP) products. For example, only RUSF and FVO have been procured from West African suppliers by WFP in recent years.
- USAID/FFP procures most SNFs from the United States, making it the procurement channel most immediately relevant to policy.

Results for scenarios 2-4 can be found in **Appendix 3**.

Table 10: Comparing Product Cost ($ Per MT) Based on Different Product Specifications and Procurement Channels.

		CSB+			CSWB			SC+			RUSF			Oil (FVO)		
Specification	rocurement channe	Detail	Cost/ MT	Data source	Detail	Cost/ MT	Data source	Detail	Cost/ MT	Data source	Detail	Cost/ MT	Data source	Detail	Cost/ MT	Data source
Standard	imported from USA	25kg bag	556 (443, 700)	USAID1	25kg bag	1000	Didion2	SC+ 1.9kg bag	1555 (1406, 1657)	USAID1	100g sachet	2,835	USAID4	4L can	1256 (1123, 1534)	USAID1
Experimental	imported from USA	1.2kg bag	1,204	Didion2	1.2kg bag	1,442	Didion2	SC+ with amylase (SC+A)	1,886	Didion2	N/A	N/A	N/A	N/A	N/A	N/A
Standard	LRP	Suppliers in Africa	642	WFP3	N/A	N/A	WFP3	SC+ Suppliers in Africa	1,407	WFP3	Suppliers in West Africa	3,107	WFP3	Suppliers in West Africa	1,487	WFP3

1 *Mean (Min, Max) of USAID 2016 -2017 historical transaction prices for products imported from USA to West Africa*, with additional in-country inspection cost provided by WFP-Sierra Leone for imported foods. * West Africa as defined by USAID Food For Peace trading route regions.*

2 *Price quotes for experiemental specifications (with initial captial cost) at scaled production of >500 MT from Didion, a major US food aid supplier (to mimic common program procurement scale), with additional in-country inspection cost provided by WFP-Sierra Leone for imported foods.*

3 *Only mean was provided by WFP based on WFP's 2016-2018 historical transaction prices for products locally or regionally procured in Africa, with additional in-country inspection cost provided by WFP-Sierra Leone for locally procured foods. All LRP products were in standard specifications. No suppliers in West Africa had produced CSB+ and SC+ for WFP in the past, so LRP cost data for CSB+ and SC+ came from suppliers outside of West Africa in the continent.*

4 *Only one historical transaction available for RUSF throughout USAID 2016 -2017 which was imported from USA to Burundi, with additional in-country inspection cost provided by WFP-Sierra Leone for imported foods.*

Table 11: Comparing International Freight Cost ($ per MT) from USA to Freetown, Sierra Leone

Flag Vessel	CSB+ & CSWB. (25 kg Bag)	SC+A (1.9 kg Bag), CSB+ & CSWB (1.2 kg Bag) 3	RUSF. (100 g)	FVO (4 L Can)
P21	130	137	145	150
P32	110.5	117.5	125.5	131.5

Data source: BKA Logistics.

Acronyms: CSB+: Corn Soy Blend Plus; CSWB: Corn Soy Whey Blend; SC+A: Super Cereal Plus with Amylase; RUSF: Ready-to-Use Supplementary Food; FVO: Fortified Vegetable Oil;

1 P2: US flag vessel from USA with a foreign flag vessel relay to final discharge port. P2 rates took an average between shipments from US Gulf and from US East Coast. P2 rates were used to estimate international freight cost in the cost and cost-effectiveness analyses of this study.

2 P3: foreign flag vessel entire ocean transport. P3 rates took an average between shipments from US Gulf and from US East Coast.

3 International freight rate for SC+ (1.9kg bag) was used as proxy for SC+A(1.9kg), CSB+ (1.2 kg), and CSWB (1.2kg) because of similarities in packaging size.

4 International freight rate for RUTF(100g) was used as proxy for RUSF(100g) because of similarities in packaging size.

b. Cost-effectiveness

As shown in **Figure 10** and **Figure 11**, there are two effectiveness measures: 1) recovery from MAM, and 2) sustained recovery at one month post-intervention. The formulas below describe how each cost-effectiveness measure was calculated:

Per Arm – Cost per Child Recovered from MAM

$$= \frac{Number\ of\ Enrolled\ Children * Cost\ per\ Enrolled\ Child}{Number\ of\ Enrolled\ Children * \%\ Recovered\ from\ MAM}$$

$$= \frac{Cost\ per\ Enrolled\ Child}{\%\ Recovered\ from\ MAM}$$

Per Arm – Cost per Child Who Sustained Recovery

$$= \frac{Number\ of\ Enrolled\ Children * Cost\ per\ Enrolled\ Child}{Number\ of\ Enrolled\ Children * \%\ Recovered\ from\ MAM * \%\ Sustained\ Recovery}$$

$$= \frac{Cost\ per\ Enrolled\ Child}{\%\ Recovered\ from\ MAM * \%\ Sustained\ Recovery}$$

Cost-effectiveness uncertainty ranges corresponding to cost per child recovered were constructed using the following process:

- Marginal mean effects of recovery and sustained recovery were estimated based on predicted probabilities of the adjusted effectiveness models (see **Section 7.2** for more details).
- Uncertainty ranges for recovery rates across arms were constructed based on 95% CIs of marginal mean effects from the adjusted model.

Cost-effectiveness uncertainty ranges corresponding to cost per child who sustained recovery were constructed using the following process:

- Depending on the specific model for sustained recovery, a recovered child who missed the one-month visit was included in the analysis by simulating his or her outcome as either all relapsed or as all sustained recovery.
- The average point estimate from the imputation simulations of the adjusted model provided point estimates of the arm-specific sustained recovery rates used to construct uncertainty ranges.

In addition to uncertainty around effectiveness, uncertainty around cost to caregivers as described in the previous section was also included in the cost-effectiveness uncertainty ranges for combined program and caregiver perspectives.

All cost and cost-effectiveness analyses were conducted using Microsoft Excel 2016 and R version 3.4.1.

9.2 Results

a. Costs – Program Perspective

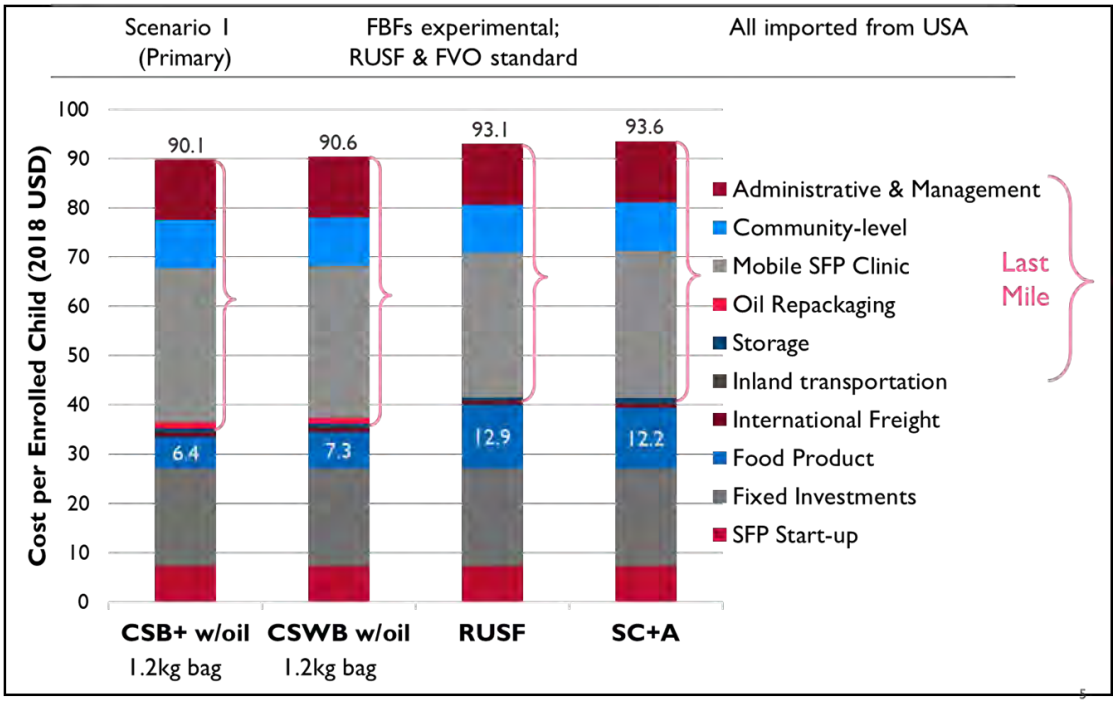
Product Cost and International Freight Cost

To mimic procurement sizes used in typical programming contexts, the price quotes for novel specifications (formulation and/or packaging) were estimated at scaled production >500 MT as shown in **Table 11**. Assuming all foods were imported from the United States, the product cost per MT was much higher for novel specifications compared to their standard counterparts. This seemed to be because some amount of initial capital cost per MT was applied to all novel specifications in the price quotes provided by Didion Milling. At a more realistic program scale, these costs would be amortized over a larger quantity, and the cost per MT would be lower. After deducting the initial capital cost from the price quotes, the absolute cost difference between the two packaging specifications (novel versus standard) was the same for both CSB+ and CSWB flours. Similarly, the absolute difference between the two was the same for packaging in 25 kg bags compared to 1.2 kg bags. For both CSB+ and CSWB, switching packaging into better material and smaller sizes increased overall product cost more than modifying the formulation.

Another arm-specific factor was the cost difference between international and locally procured foods. Product cost per MT was higher when a standard product was locally procured than when imported from the United States (**Table 10**). Of the four foods, RUSF had the most comparable local procurement data obtainable, resulting in a reported product cost difference (local versus international) of $272/MT. There were two elements driving this higher cost: higher local procurement costs and higher in-country inspection costs. The mean procurement price for RUSF paid by WFP to West African suppliers from 2016 to 2018 was approximately $100/MT higher than RUSF procured by USAID from American suppliers from 2016 to 2017. This is despite the U.S.-based commodity price including inspection costs from a U.S.-based lab. Inspection costs of locally procured RUSF were $1,630 per tranche in addition to the product price. Products imported from the United States only incurred additional inspection fees of $25 per shipping container. A major USAID freight forwarder provided international freight quotes for standard products on two types of flag vessels. These quotes served as proxies for the cost of shipping novel products with similar packaging specifications from the United States to Sierra Leone. Shipping via U.S. flag vessels was approximately $20/MT more expensive than via foreign vessels (**Table 11**). Comparing across the study foods' different specifications, packaging differences did not substantially affect international freight cost per MT.

Cost from the Program Perspective

Figure 12: Program Perspective Cost per Enrolled Child.



The primary scenario analysis considers all foods with the product specifications used in the study and imported from the United States. Cost per enrolled child in this scenario was similar across the four arms, with CSB+ w/oil being the lowest at $90 per child and SC+A the highest at $94 per child (**Figure 12**).

On average, children collected 3.3 to 3.4 rations during treatment in each arm. The largest *difference* in cost across arms was product cost. Other minor sources of cost difference included oil repackaging cost (~$1 per child for CSB+ w/oil and CSWB w/oil only) and supply chain costs per enrolled child including international freight, inland transportation, and storage (~$0.60 per child difference between the lowest [RUSF] and highest [SC+A] arms).

The CSB+ w/oil and CSWB w/oil arms had about $2 per child higher mobile SFP clinic-attributable costs than RUSF and about $1 per child higher than SC+A. Food-specific differences in the amount of time spent at clinic arising from the cooking demonstration and oil distribution likely account for these differences in clinic-attributable costs (see "last mile" section of **Figure 12**).

Though product cost drove cost *differences* across arms, product cost was not the largest *component* of the total cost. The largest component of total cost in this context was the cost associated with the mobile SFP clinic (see **Section 3.1** for more details). The mobile-based SFP may have had higher transportation costs than site-based or integrated models due largely to the biweekly Land Cruiser usage.

The mobile SFP clinic cost is also part of the last mile costs, defined as in-country costs including inland transportation, storage, oil repackaging, distribution/clinic, community-level activities, and administrative and management activities (**Figure 12**). Last mile costs constituted more than half of the total cost

across arms, contrasting with FAQR findings in another report about last mile costs in other types of food aid programs (e.g., blanket supplementary feeding).[33] Additionally, lower-than-expected caseload in the catchment areas increased the last mile costs per enrolled child because full capacity of input resources was not reached.

b. Cost-effectiveness – Program Perspective

The cost-effectiveness results are shown for cost per child recovered from MAM in **Figure 13** and for cost per child with sustained recovery in **Figure 14**. From the program perspective, CSB+ w/oil had the most cost-effective point estimate for both recovery and sustained recovery. For cost per child recovered from MAM, all arms were similarly cost-effective, with mostly overlapping uncertainty ranges. While CSB+ w/oil was only $2 less than RUSF in cost *per child recovered from MAM*, CSB+ w/oil was $18 less than RUSF in cost *per child with sustained recovery,* with barely overlapping uncertainty ranges. This suggests that RUSF was significantly less cost-effective than CSB+ when considering cost-effectiveness based on sustained recovery.

Figure 13: Cost per Child Recovered from Moderate Acute Malnutrition (MAM): Program Perspective Versus Program & Caregiver Perspective

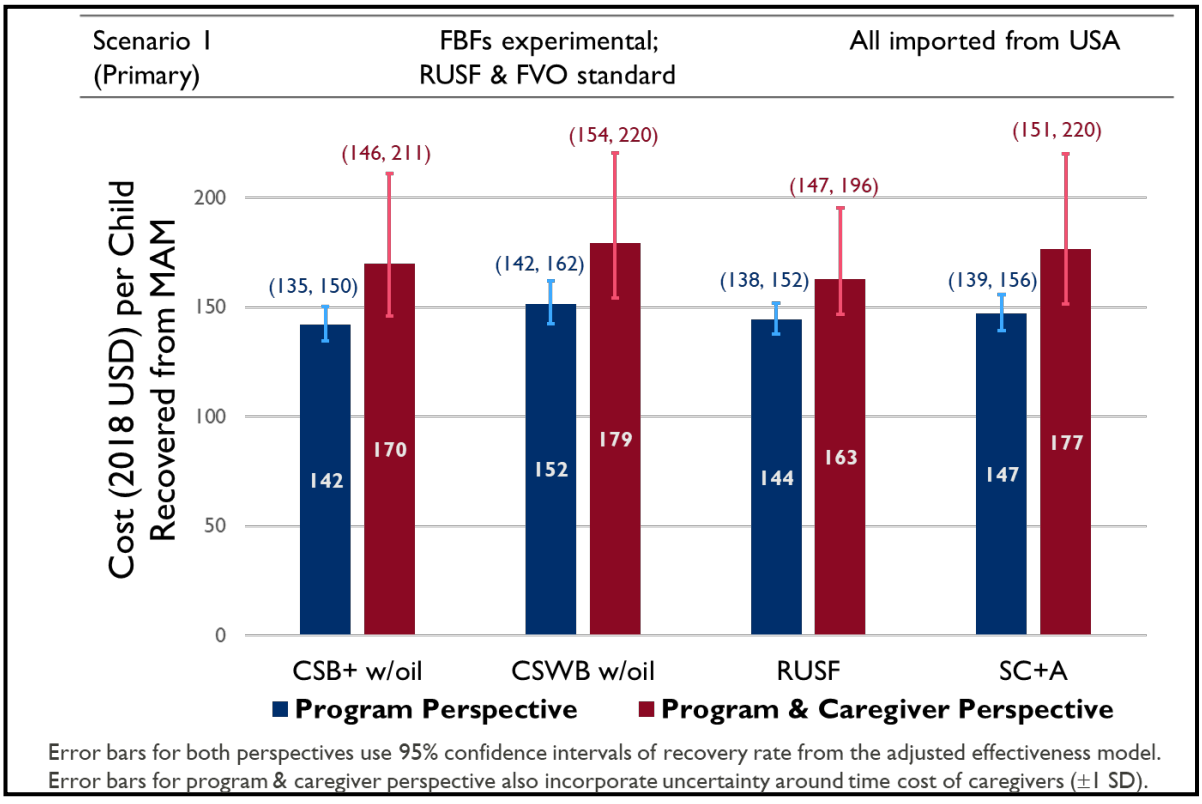
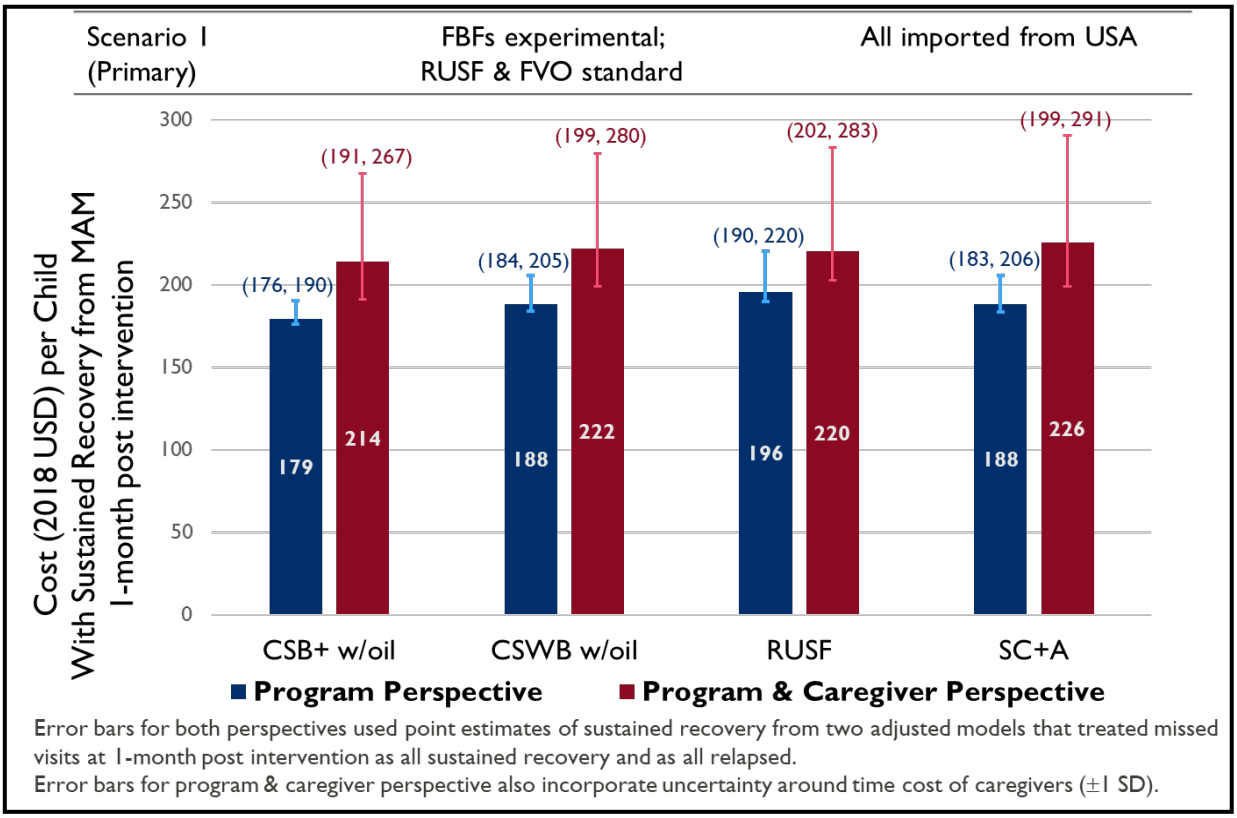


Figure 14: Cost per Child Sustained Recovery 1-month post intervention: Program Perspective Versus Program & Caregiver Perspective



Error bars for both perspectives used point estimates of sustained recovery from two adjusted models that treated missed visits at 1-month post intervention as all sustained recovery and as all relapsed.
Error bars for program & caregiver perspective also incorporate uncertainty around time cost of caregivers (±1 SD).

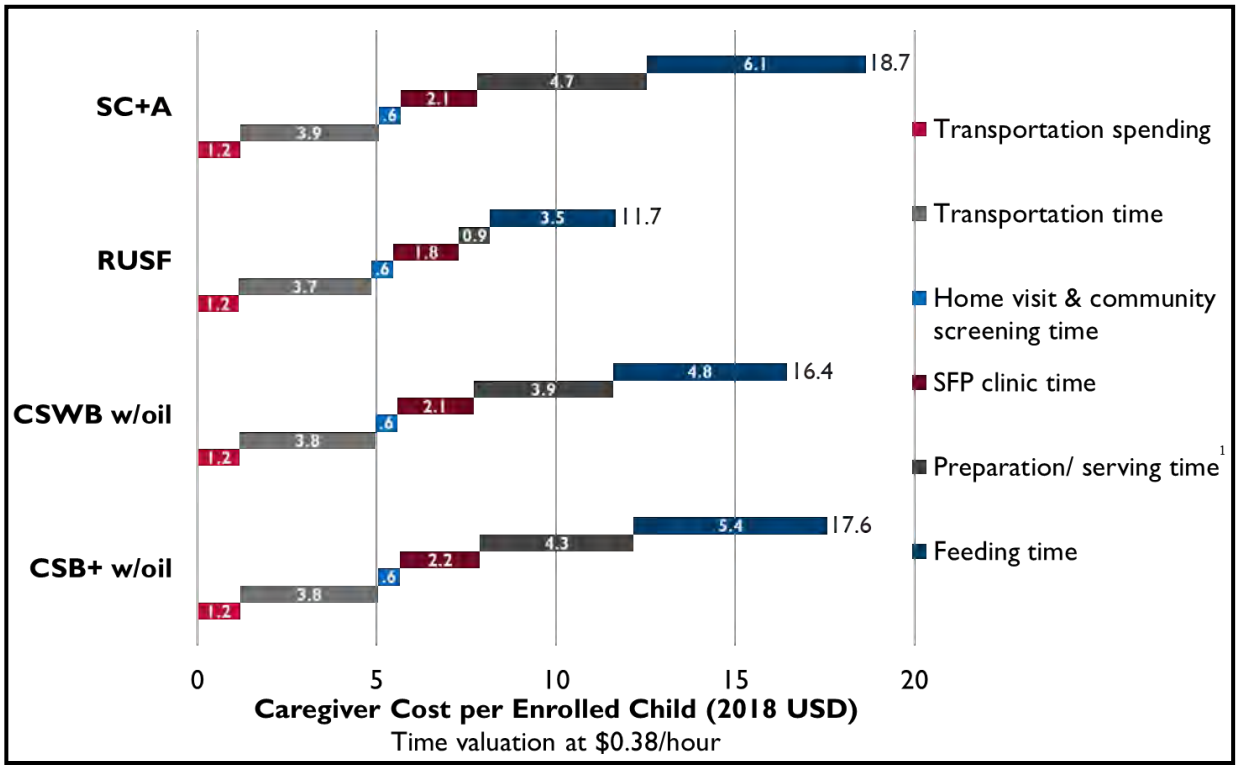
c. Costs – Caregiver Perspective

Total cost per enrolled child from the caregiver perspective varied across arms. As the only ready-to-use product of the four SNFs, RUSF had the lowest total cost for caregivers: about $5-7 per enrolled child lower than the FBF arms (**Figure 14**), though this figure is dependent on the valuation of time chosen for the analysis.

From the caregiver perspective, across-arm cost differences resulted from differences in time that caregivers devoted to relevant activities both during the SFP and in the home. Among these activities, the largest driver of cost differences was the observed caregivers' time spent on study food preparation/serving and feeding at home. Even though 32% of caregivers in the RUSF arm were observed to use RUSF as an ingredient to cook a home meal for the children, the average cost attributable to time over the entire sample in the RUSF arm remained lower than in the FBF arms. Caregivers in the RUSF arm also spent less time feeding the beneficiary child than in the FBF arms. Among the three FBF arms, caregivers receiving SC+A consistently spent the most time preparing and feeding per meal.

The opportunity cost of caregivers' time spent at the SFP clinic was a minor source of cost differences across arms, which was about $0.30 per enrolled child lower in the RUSF arm. During the first visit, where children were enrolled into the program, caregivers in the RUSF arm spent about 36-53 minutes less at

Figure 15: Caregiver Perspective Cost per Enrolled Child



the SFP clinic than those receiving any of the FBFs. This difference occurred because the individual taste test for RUSF took less time than the group cooking demonstration for FBFs at the first/enrollment visit.

Relating to food-specific differences in caregivers' opportunity costs at the intake visit, one theme identified during focus group discussion was "forgetfulness" due to factors such as the time-intensive nature of the demonstration and a generalized lack of attention as a result. This is described in more detail in **Section 9.2.b**.

Caregivers' monetary and time costs for transportation and community-level activities were not arm-specific (**Figure 15**). About 16 to 17% of caregivers who completed an in-depth interview (IDI) reported having to pay on average $1.10 for transportation each way to collect a single biweekly ration of study foods. Therefore, the average transportation spending for the whole sample was about $0.40 per enrolled child. Regarding opportunity costs related to transportation time, caregivers reported traveling on average of three hours round-trip to collect a biweekly ration of study foods, which totaled about $3.80 per enrolled child. The opportunity cost of caregivers' time spent in community-level activities, including home visits and community screening, was $0.60 per enrolled child. Other transportation and travel challenges included the long distance to get to the clinic and environmental obstacles such as rain, flooding, rough terrain, and hot sun. Travel by foot was the only mode of transportation reported across all arms. This was confirmed during FGDs; those who walked to the clinic explained that their personal limited resources were a constraint to paying for transportation, which resulted in spending the majority of the day getting to and from the clinic.

> *I walked with my naked foot to come. If you ask anybody to bring you, to leave his own business, he will ask you to pay big money, and my husband is bankrupt; even me, I am hard-up. What I do is walk with my naked foot. (FGD #7, Respondent 6)*

To assess variability in caregivers' time, uncertainty ranges for total cost per enrolled child were estimated from the caregiver perspective (**Table 12**). The values for the upper and lower bounds of the RUSF uncertainty ranges were consistently smaller than the corresponding bounds of uncertainty ranges for the FBF arms. However, due to the large variation within each arm, the uncertainty ranges for RUSF still overlapped with those for the FBF arms.

d. Cost-effectiveness - Cost from All Stakeholder Perspectives

Different sets of stakeholders contributed to the cost of the SFP in several components. **Table 12** shows the summarized cost per enrolled child from each of the five perspectives defined in **Figure 9**. Cost per enrolled child from the donor and implementer perspective was the highest, regardless of the arm. Meanwhile, cost per enrolled child from the caregiver perspective had the largest differences across arms. Opportunity cost of lead mother volunteers and free space provided by the government was $7 per child for each arm.

Comparing the two combined perspectives (program versus program and caregiver) in **Table 12**, SC+A remained the most expensive arm in both. While RUSF was about $3 per child more than CSB+ w/oil and CSWB w/oil from the program perspective, RUSF was about $3 per child less than these two arms when the program and caregiver perspectives were combined.

e. Cost-effectiveness – Combined Program and Caregiver Perspective

The cost-effectiveness results are shown in **Figure 13** for cost per child recovered and in **Figure 14** for cost per child with sustained recovery.

From the combined program and caregiver perspective, RUSF had the most cost-effective point estimate for cost per child recovered from MAM, while CSB+ w/oil had the most cost-effective point estimate for cost per child with sustained recovery. From the combined perspective, all uncertainty ranges for the two cost-effectiveness measures (program and caregiver) overlapped across all four arms.

9.3 Summary

Table 12: Cost per Enrolled Child Corresponding to Each Perspective

Perspective	CSB+ w/oil	CSWB w/oil	RUSF	SC+A
Donor & Implementer	83	83.5	86	86.5
Government & Volunteer[1]	7.1	7.1	7.1	7.1
Caregiver	17.6	16.4	11.7	18.7
	(7.6, 36.4)[2]	(7.5, 32.7)[2]	(6.1, 26.7)[2]	(8.0, 38.6)[2]
Program	**90.1**	**90.6**	**93.1**	**93.6**
Program & Caregiver	**107.7**	**107**	**104.7**	**112.3**
	(97.7, 126.5)[3]	**(98.1, 123.3)[3]**	**(99.2, 119.8)[3]**	**(101.6, 132.2)[3]**

[1] Includes storage space provided by government, community health worker Base Pay by government, and volunteer opportunity cost of lead mothers involved in community-level activities.

[2] Constructed based on mean ± 1 SD of time per occasion for caregiver activities that were sources of across arm differences

[3] Constructed by incorporating the ranges for caregiver perspective cost

From the program perspective, cost per enrolled child and cost per child recovered from MAM were similar across the four foods when they were imported from the United States. Cost per child with sustained recovery was lowest for the CSB+ w/oil arm, and RUSF was less cost-effective than CSB+ w/oil.

Cost per enrolled child from the caregiver perspective was lower for the RUSF arm mostly due to its ready-to-use nature. However, the addition of this perspective to the analysis did not alter conclusions that the four foods performed similarly in terms of cost per recovered child. CSB+ w/oil remained lower in cost per child with sustained recovery with overlapping uncertainty ranges.

The only out-of-pocket spending for caregivers was about $0.40 per enrolled child on average to pay for transportation. The majority of the caregiver costs was the opportunity cost of caregivers' time spent in relevant activities such as transportation, SBCC, SFP clinic, and study food preparation/serving and feeding at home. The largest drivers of caregiver cost differences across arms were preparation/serving and feeding, especially between RUSF and the three FBFs. Among the three FBFs, SC+A was observed to have the highest preparation and feeding time. Differences in total caregiver perspective cost per enrolled child was $7 between the lowest cost option, RUSF, and the highest cost option, SC+A.

10. Other Factors Influencing Effectiveness

10.1 Methods

Quantitative outcomes related to factors influencing the effectiveness of the four foods were based on data from the subsamples of households who underwent in-home observation (IHO) for five consecutive days and/or participated in an IDI and linked with the clinic data for outcomes. Qualitative information was from either enumerator notes recorded during in-home observations or from FGDs.

a. Consumption and Sharing

The following descriptive statistics were calculated: observed consumption of the study food ration by the target child and others, and observed sharing of the study food ration, by both study arm and day of observation; self-reported sharing of the study food ration and reasons for sharing; self-reported dietary diversity. In addition, displacement of household complementary foods was assessed through a 24-hour dietary diversity recall using bivariate logistic regressions. Food groups consumed by children were reported by caregivers based on a 24-hour reference period, determined based on the IYCF minimum diet diversity (MDD) score, and regressed on assigned study arm.

b. Recipe Adherence Behavior

All caregivers received counseling regarding the correct quantities and proportions of water, flour, and, where relevant, FVO (see **Figure 1**). The unit of measurement was a standard baby feeding cup that was readily available in every village and most households (see **Section 5.2**). In all households surveyed, this common unit was available.

The recommendations for the quantity of FVO and FBF were made in grams (g). However, the unit of measurement at the household and for data collection was milliliters (ml). The target quantities of flour for the FBF arms were therefore 85.7 g = 150 ml for CSB+ w/oil; 85.7 g = 150 ml for CSWB w/oil; and 135.7 g = 210 ml for SC+A. Caregivers were categorized as reporting the "correct" amount of flour if they were within ±10 ml of the target amount. In the CSWB w/oil and CSB+ w/oil arms, the target quantity of FVO was 26 g = 30 ml. Caregivers were categorized as reporting the correct amount of FVO if they were within ± 5 ml of the target amount.

The target ratio of flour to water in the CSB+ w/oil and the CSWB w/oil arms was 0.33 and 0.4 for SC+A. The same lower and upper bounds used to categorize correct flour measures were applied to water (450 ml ± 10 ml) in the CSB+ w/oil and CSWB w/oil arms and in the SC+A arm (525 ml ± 10 ml). The upper bound of the correct flour to water ratio was therefore set at 160/440 (~0.36) and the lower bound at 140/460 (~0.30) in the CSB+ w/oil and the CSWB w/oil arms. In the SC+A arm, the upper bound was set at 200/535 (~0.43) and the lower bound at 220/515 (~0.37).

Care and feeding practices observed during IHOs were compared to self-reported information in the IDIs. Following guidance correctly was defined as when caregivers reported using the correct ingredients and no other ingredients. For RUSF, the definition of adherence was defined as eating it straight out of the packet and not mixing it with other foods. Caregivers observed cooking the FBFs using only the recommended ingredients were categorized as correctly adhering to the recipe guidance. Similarly, caregivers observed feeding the beneficiary child the RUSF without mixing it with other foods were categorized as adhering to the recipe guidance.

Descriptive summary data reported the percentage of respondents self-reporting or the percentage of households in which the behavior was observed. Descriptive statistics included the percentage of respondents *reporting* to use of correct ingredients compared to percentage of households *observed* to use the correct ingredients at least once during the five-day period; the percentage of respondents accurately reporting the quantity of ingredients used at last preparation for the FBF arms; and the percentage of households observed to measure all the ingredients at least once during the five-day observation.

10.2 Results

a. Demographic Characteristics of Interviewed and Observed Households

In the subsample, a total of 962 caregivers were interviewed (IDI), of whom 321 households were also observed (IHO). Demographic differences between the two subsamples are detailed in **Table 13**: caregivers were generally married, uneducated, and in their late 20s across arms and subsamples. Beneficiary children were generally still breastfeeding (>75%) and had similar MUAC at enrollment (12 cm ± 0.3) and similar graduation rates, illustrating that the subsample was comparable to the full study sample.

Consumption and Sharing

Figure 16: Observed consumption of the study food at least once during the 5-day observation period

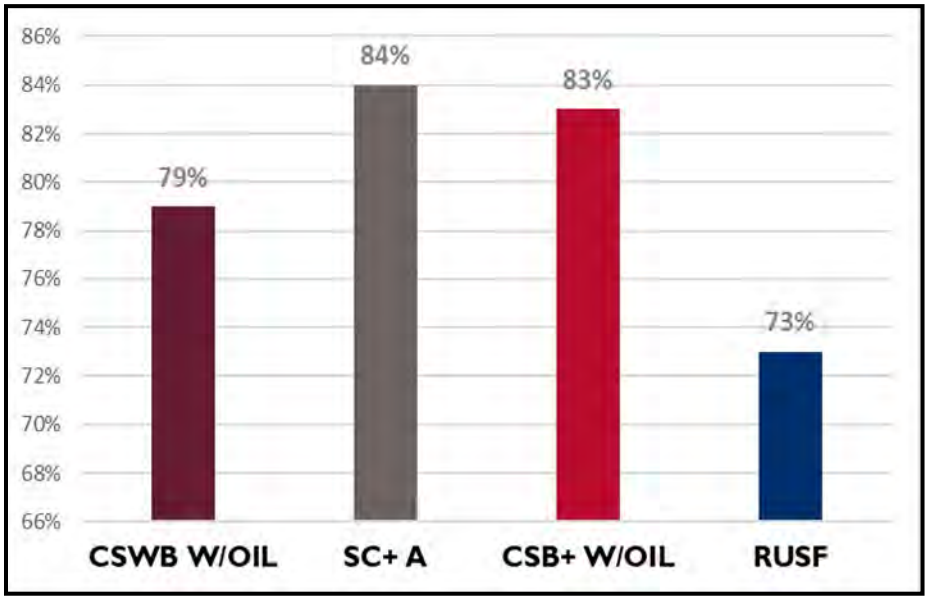
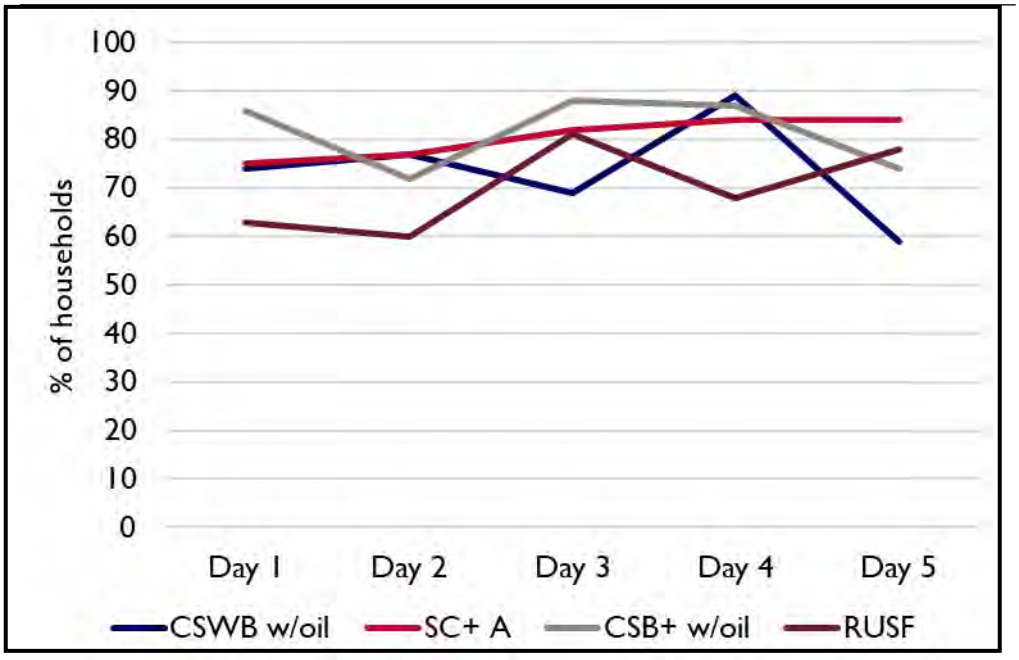


Figure 17: Observed consumption by study arm and day of observation



Overall, the target recipients were observed consuming the study foods at least once during the five-day observation period in 80% of observed households. **Figure 16** shows the percentage of households in which the target child was observed eating the study food at least once during the five-day observation period, by study arm. There were no discernable consumption trends by day of observation **Figure 17**.

Consumption was also discussed among caregivers in FGDs. Overall and within each arm, nearly all of the respondents reported that the beneficiary child consumed the ration normally. When asked about daily ration consumption, respondents across arms reported serving the study food more than once per day, with some in the CSB+ w/oil and RUSF arms reporting serving the food only once. Having leftovers was evident in all FBF arms, with the largest number in the CSB+ w/oil arm. Leftovers were not mentioned in the RUSF arm during FGDs.

> *She eats it [the SC+A] three times, but I cook it only one time. I only cook the one that will be enough for her. (FGD #8, Respondent 6)*

Other consumption and utilization patterns emerged during FGDs. In the RUSF and CSWB w/oil arms, the ration was consumed by caregivers. Respondents in the RUSF arm reported tasting or consuming the food during feeding times. "Testing" the food prior to giving it to children was a perceived common behavior. Consumption of the food by the caregivers occurred in the CSWB w/oil arm when children refused to consume the entire portion and leftovers remained. No respondents indicated that their consumption of the study affected the 14-day ration supply.

> *I also tasted it [the RUSF], because as a mother whenever you are feeding a child you have to taste it first. (FGD #9, Respondent 5)*

> *For me, I give the food [CSWB w/oil] to my child and after she has eaten a little, I take the rest [of the food] in my own pan and we go to the farm. (FGD #11, Respondent 6)*

Giving the FBFs foods to the beneficiary children raw, without cooking it, was a theme that arose among the SC+A and CSB+ w/oil study arms. This was expressed as an act of appeasing a crying child by one caregiver in the CSB+ w/oil arm.

> *My child started eating the raw/dry blended [CSB+] when I received the last supply. she eats it [raw] once a day, only when she cries for it while I am preparing to cook it [the food]. (FGD #7, Respondent 1)*

Table 13: Demographic Characteristics of Respondents and Observed Households

	CSWB		SC+		CSB+		RUSF		Total	
	IDI	IHO	IDI	IHO	IDI	IHO	IDI	IHO	IDI	IHO
Child enrollment characteristics	*n (%) or mean ± SD*		*n (%) or mean ± SD*		*n (%) or mean ± SD*		*n (%) or mean ± SD*		*n (%) or mean ± SD*	
n	**213 (22.14%)**	**69 (21.5%)**	**249 (25.88%)**	**83 (25.86%)**	**251 (26.09%)**	**90 (28.04%)**	**249 (25.88%)**	**79 (24.61%)**	**962**	**321**
Male gender	82 (38.5%)	22 (31.9%)	96 (38.6%)	33 (39.8%)	108 (43%)	37 (41.1%)	116 (46.6%)	39 (49.4%)	402 (41.8%)	131 (40.8%)
Age, months	12.5 ± 8.1	11.9 ± 6.1	13.1 ± 8.7	12.6 ± 8	12.6 ± 7.8	11.5 ± 5.9	14 ± 8.9	13.1 ± 7.3	13.1 ± 8.4	12.3 ± 6.9
Breastfeeding	172 (80.8%)	55 (79.7%)	195 (78.3%)	65 (78.3%)	205 (81.7%)	77 (85.6%)	188 (75.5%)	58 (73.4%)	760 (79%)	255 (79.4%)
Transferred from SAM	40 (18.8%)	15 (21.7%)	66 (26.5%)	22 (26.5%)	62 (24.7%)	25 (27.8%)	71 (28.5%)	22 (27.8%)	239 (24.8%)	84 (26.2%)
Twin (at enrollment)	9 (4.2%)	3 (4.3%)	10 (4%)	3 (3.6%)	10 (4%)	5 (5.6%)	0 (0%)	0 (0%)	29 (3%)	11 (3.4%)
Admitted to hospital (past 2 weeks)	2 (0.9%)	1 (1.4%)	9 (3.6%)	0 (0%)	10 (4%)	4 (4.4%)	8 (3.2%)	2 (2.5%)	29 (3%)	7 (2.2%)
Morbidities (presence of in past 7 days)										
Fever	76 (35.7%)	17 (24.6%)	83 (33.3%)	21 (25.3%)	72 (28.6%)	16 (17.6%)	66 (26.5%)	20 (25.3%)	297 (30.8%)	74 (23%)
Diarrhea	19 (8.9%)	5 (7.2%)	18 (7.2%)	3 (3.6%)	15 (6%)	4 (4.4%)	11 (4.4%)	4 (5.1%)	63 (6.5%)	16 (5%)
Vomit	13 (6.1%)	4 (5.8%)	16 (6.4%)	2 (2.4%)	13 (5.2%)	3 (3.3%)	11 (4.4%)	2 (2.5%)	53 (5.5%)	11 (3.4%)
Cough	57 (26.8%)	18 (26.1%)	59 (23.7%)	14 (16.9%)	56 (22.2%)	11 (12.1%)	61 (24.5%)	18 (22.8%)	233 (24.2%)	61 (18.9%)
Average weight (kg)	6.6 ± 1	6.6 ± 1	6.6 ± 1.1	6.6 ± 1	6.6 ± 1	6.5 ± 1	6.7 ± 1.1	6.6 ± 1	6.6 ± 1	6.6 ± 0
Average length (cm)	67.4 ± 5.9	67.4 ± 5.7	67.8 ± 6.6	67.2 ± 6	67.6 ± 6	67 ± 5.4	68.1 ± 6.4	67.8 ± 6.1	67.7 ± 6.2	67.3 ± 5.8
Average MUAC (cm)	12 ± 0.3	12 ± 0.3	11.9 ± 0.3	11.9 ± 0.2	12 ± 0.3	12 ± 0.3	12 ± 0.3	11.9 ± 0.3	12 ± 0.3	12 ± 0.3
Weight-for-age z-score	-2.8 ± 0.8	-2.7 ± 0.8	-2.8 ± 0.7	-2.8 ± 0.7	-2.8 ± 1	-2.7 ± 0.7	-3 ± 0.9	-3 ± 0.8	-2.8 ± 0.9	-2.8 ± 0.8
Length-for-age z-score	-2.7 ± 1.2	-2.5 ± 1.3	-2.7 ± 1.2	-2.8 ± 1.2	-2.7 ± 1.2	-2.6 ± 1.2	-2.7 ± 1.2	-2.9 ± 1.3	-2.7 ± 1.2	-2.7 ± 1.3
Weight-for-length z-score	-1.7 ± 0.7	-1.6 ± 0.7	-1.7 ± 0.7	-1.6 ± 0.7	-1.7 ± 0.7	-1.6 ± 0.7	-1.7 ± 0.7	-1.6 ± 0.7	-1.7 ± 0.7	-1.6 ± 0.7
Caregiver & household characteristics										
Caregiver's age, years	28.6 ± 8.9	28.5 ± 8.1	26.8 ± 6.8	26.1 ± 6.3	29.2 ± 8.7	28.1 ± 6.8	28.1 ± 8.9	27.4 ± 8.1	28.2 ± 8.4	27.5 ± 7.3
Marital status										
Married	183 (85.9%)	61 (88.4%)	205 (82.3%)	70 (84.3%)	219 (87.3%)	79 (87.8%)	217 (87.1%)	67 (84.8%)	824 (85.7%)	277 (86.3%)
Separated	2 (0.9%)	0 (0%)	5 (2%)	1 (1.2%)	4 (1.6%)	1 (1.1%)	3 (1.2%)	0 (0%)	14 (1.5%)	2 (0.6%)
Single	28 (13.1%)	8 (11.6%)	37 (14.9%)	10 (12%)	28 (11.2%)	10 (11.1%)	29 (11.6%)	12 (15.2%)	122 (12.7%)	40 (12.5%)
Level of education										
None	122 (57.3%)	40 (58%)	130 (52.2%)	42 (50.6%)	147 (58.6%)	51 (56.7%)	136 (54.6%)	39 (49.4%)	535 (55.6%)	172 (53.6%)
Some or completed primary	47 (22%)	15 (21.7%)	57 (22.9%)	16 (19.3%)	56 (22.3%)	23 (25.5%)	62 (24.9%)	22 (27.9%)	222 (23.1%)	76 (23.7%)
Some, completed, or more than second	44 (20.6%)	14 (20.2%)	61 (24.5%)	24 (28.9%)	48 (19.1%)	16 (17.8%)	51 (20.5%)	18 (22.8%)	204 (21.2%)	72 (22.4%)
1 or more Females age 65+	27 (12.7%)	7 (10.1%)	36 (14.4%)	11 (13.2%)	29 (11.6%)	12 (13.3%)	26 (10.4%)	13 (16.5%)	118 (12.2%)	43 (13.4%)
HFIAS category										
Food Secure	51 (23.9%)	18 (26.1%)	63 (25.3%)	27 (32.5%)	78 (31.2%)	37 (41.1%)	67 (26.9%)	28 (35.4%)	259 (27%)	110 (34.3%)
Mildly Food Insecure Access	0 (0%)		1 (0.4%)		0 (0%)		1 (0.4%)		2 (0.2%)	
Moderately Food Insecure Access	27 (12.7%)	10 (14.5%)	37 (14.9%)	12 (14.5%)	21 (8.4%)	11 (12.2%)	30 (12%)	9 (11.4%)	115 (12%)	42 (13.1%)
Severely Food Insecure Access	135 (63.4%)	41 (59.4%)	148 (59.4%)	44 (53%)	151 (60.4%)	42 (46.7%)	151 (60.6%)	42 (53.2%)	585 (60.9%)	169 (52.6%)
SES quintiles										
Lowest	50 (23.7%)	16 (23.5%)	52 (21.3%)	20 (24.7%)	60 (24.2%)	26 (28.9%)	53 (21.4%)	14 (17.7%)	215 (22.6%)	76 (23.9%)
Mid-Low	44 (20.9%)	16 (23.5%)	36 (14.8%)	12 (14.8%)	48 (19.4%)	16 (17.8%)	57 (23%)	19 (24.1%)	185 (19.5%)	63 (19.8%)
Medium	39 (18.5%)	11 (16.2%)	49 (20.1%)	16 (19.8%)	47 (19%)	15 (16.7%)	57 (23%)	20 (25.3%)	192 (20.2%)	62 (19.5%)
Mid-High	43 (20.4%)	14 (20.6%)	52 (21.3%)	19 (23.5%)	42 (16.9%)	14 (15.6%)	44 (17.7%)	12 (15.2%)	181 (19%)	59 (18.6%)
Highest	35 (16.6%)	11 (16.2%)	55 (22.5%)	14 (17.3%)	51 (20.6%)	19 (21.1%)	37 (14.9%)	14 (17.7%)	178 (18.7%)	58 (18.2%)
Outcomes										
Graduate	143 (67.1%)	44 (63.8%)	163 (65.7%)	52 (62.7%)	177 (70.2%)	58 (63.7%)	164 (65.9%)	54 (68.4%)	647 (67.3%)	208 (64.6%)
Developed SAM	30 (14.1%)	11 (15.9%)	33 (13.3%)	14 (16.9%)	31 (12.3%)	10 (11%)	37 (14.9%)	11 (13.9%)	131 (13.6%)	46 (14.3%)
Default	6 (2.8%)	1 (1.4%)	9 (3.6%)	5 (6%)	4 (1.6%)	2 (2.2%)	13 (5.2%)	5 (6.3%)	32 (3.3%)	13 (4%)
Death	0 (0%)	0 (0%)	1 (0.4%)	0 (0%)	0 (0%)	0 (0%)	3 (1.2%)	1 (1.3%)	4 (0.4%)	1 (0.3%)
Fail	34 (16%)	13 (18.8%)	42 (16.9%)	12 (14.5%)	40 (15.9%)	21 (23.1%)	32 (12.9%)	8 (10.1%)	148 (15.4%)	54 (16.8%)

Explanations among the SC+A arm were offered, which included preventing sickness or vomiting, while others said that the child did not like the cooked pap and preferred it raw.

> *When I am cooking it, he [the child] may be eager to eat it [SC+A] and I give him the raw one, it does not harm him or give him sickness. (FGD #6, Respondent 4)*

> *She loves the dry/raw [SC+A] more than the pap or the cooked one. (FGD #8, Respondent 3)*

In all of the FGDs, there were no mentions of mixing the raw food with any other ingredients, and some expressed that they served the food to the child alone.

> *I came to a point where I no longer cooked the blended. I added a little oil on the raw/dry blended [CSB+ flour] and mix it in a cup to give to my child to eat it alone. (FGD #7, Respondent 1)*

> *In the morning I just put some raw flour [SC+A] in the cup and water and she will eat it all…she just eats the raw flour [SC+A]. (FGD #4, Respondent 1)*

Sharing the study foods with those other than the intended recipient was both observed and reported in all study arms. Information on observed and self-reported sharing of the ration is presented in **Figure 18** by study arm. Although there were no major differences among study arms in reported sharing, observed sharing was highest in the RUSF arm and lowest in the SC+A arm. Unsurprisingly, self-reports of sharing were lower than direct observations.

Figure 18: Self-reported vs observed sharing of the study food by study arm

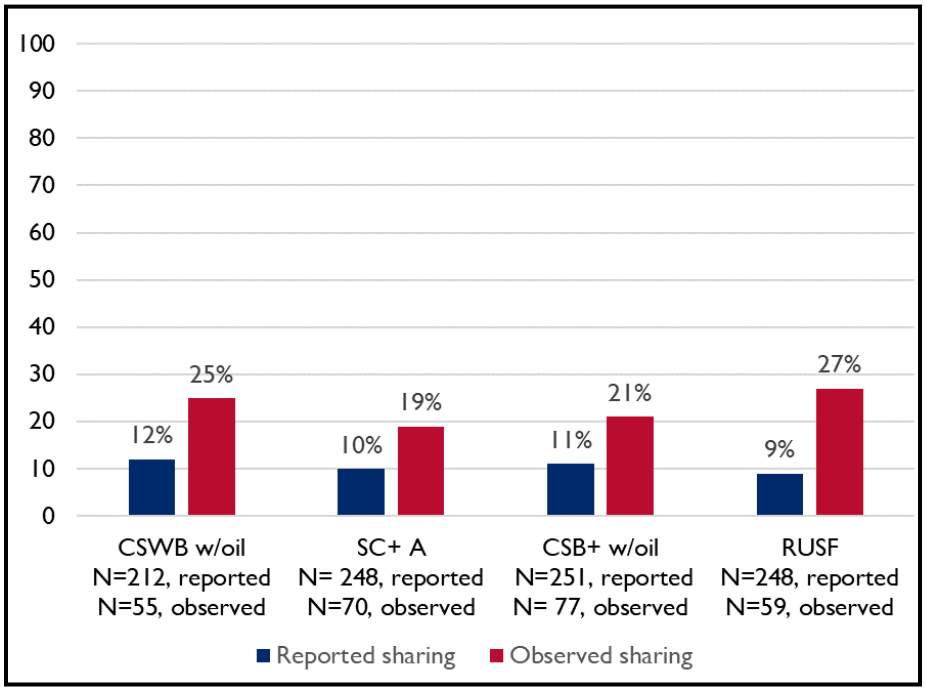


Figure 19: Observed sharing by study arm and day of observation

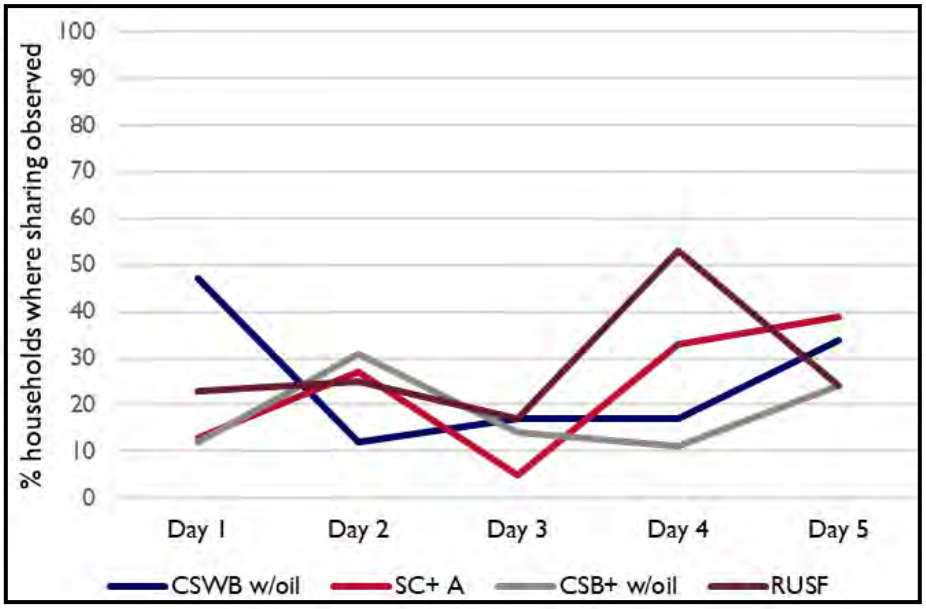


Table 14: Observed sharing by study arm and day of observation

Reported Reason	N (%)
It is good for health	330 (34)
It treats illness	291 (30)
Other children need or want	263 (27)
Other adults need or want	97 (10)
Other reason	53 (6)
Caregiver is breastfeeding the child	47 (5)
Not enough other food in the house	25 (3)
Receive more than child needs	18 (2)
Morally obligated to share	20 (2)
Not enough money to buy food	11 (1)

No discernable trends were seen in observed sharing by day of observation **Figure 19**. Additional descriptive statistics support the finding that sharing is common: 56% of respondents across all four study arms reported running out of the ration before receiving the next distribution. Very few people reported giving away (1%) or selling (0.21%) the ration to people outside of their households. Reported reasons for sharing the study food are shown in **Table 14**.

By outcome, there were few differences in observed sharing of the study foods, with less sharing observed among those who failed to graduate; however, there were differences in observed consumption by outcome. Those who graduated were observed consuming the food the most often (82% of households observed consuming the food at least once over the five-day observation), and those who failed were observed consuming the food least often (46% of households observed consuming the food at least once over the five-day observation), but this was not tested for statistical significance. These findings are summarized in **Figure 20**.

Figure 20: Percent of households observed consuming food at least one day in five by outcome

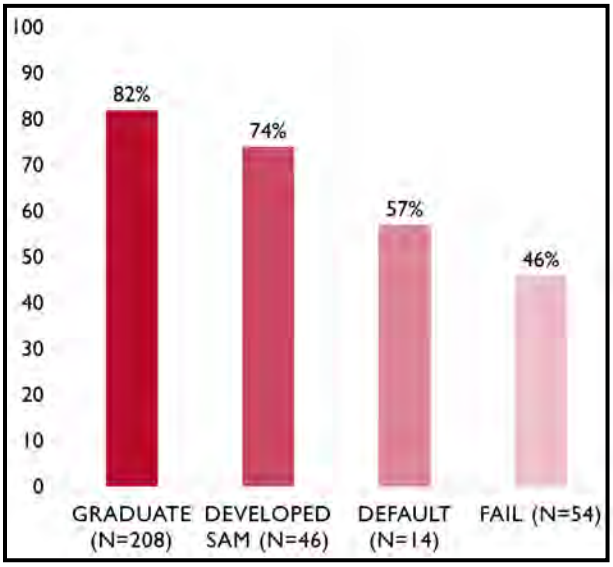
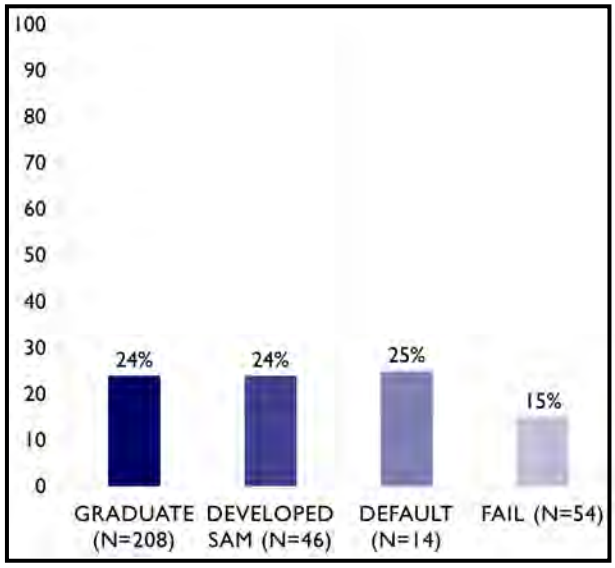


Figure 21: Percent of households observed sharing food at least one day in five by outcome



When prompted about sharing during FGDs, caregivers across all arms expressed that they did not sell or share the food outside the household, even when under social pressure from community members. Caregivers in the FBF arms reportedly shared the food with nonbeneficiary children. The sharing most often occurred with siblings or other children in the household. One caregiver went on to explain that instead of throwing away leftovers, she gave the food to the other child living in the house.

> *[Will not share] unless [for] my sister's younger child because maybe when she eats it [beneficiary child] and leave leftover, she [sister's child] will eat it. So she is the only one. (FGD #14, Respondent 1)*

Another commented:

> *I have [another] kid that does ask for the pot and I give it to her. She will eat the pap [porridge] in it.* (FGD #4, Respondent 5)

When prompted about sharing outside of the house, no respondents reported selling or exchanging their ration for other goods or services. Participants in the CSB+ w/oil arm indicated a sense of moral obligation to share with others in the community, specifically to thin children or pregnant women who had not received a ration (see **Figure 1**).

> *Besides malnourished children, I don't remember of giving this food to other people besides malnourished children and pregnant women that have skinny children in their wombs. (FGD #15, Respondent 4)*

> *If a colleague mothers' child is thin in structure and not actually the beneficiary child, if the mother request by saying please, let me have some of the blended to give to my child I will measure some of the blended and give it to her for her to cook it for her child. (FGD #3, Respondent 3)*

One respondent receiving CSB+ w/oil expressed willingness to buy the food if she saw it being sold because her child loved the food and cried for it.

> *Even my child now, when she sees the food with a colleague mothers' child, she will cry for the food and if I see where they are selling the food, I will buy it. (FGD #2, Respondent 3)*

Overall and within each arm, almost all caregivers reported not sharing because they viewed the food as "only for the beneficiary child" and pertinent to the child's development. In the CSWB w/oil and SC+A arms, sharing the food with other children was not considered acceptable—that sharing the food could transfer the "illness" to the other children.

> *The leftover [SC+A food], you should not give it to others to eat because if you do, the sickness that the beneficiary child has will transfer to the other children. (FGD #4, Respondent 7)*

> *The sickness that this child has, the other child doesn't have it, if you give the other child it [the CSWB w/oil food], it will transfer to him. That is why I don't give to other child except this child (FGD #11, Respondent 3)*

In one CSWB w/oil FGD, one caregiver stated that she did not share because she knew someone was observing them. Others in the CSB+ w/oil arm expressed they did not share because this behavior was counseled against in the photos and through scenarios used during the counseling sessions at the SFP clinic. Embedded in these descriptions were instances of knowledge translation regarding the food handling instructions, such as "we were told not to [share or sell the study food]."

> *The food is for her [beneficiary child] alone, they told me not to give to others. (FGD #4, Respondent 2)*

> *For me I don't have another thinking on this, just as they have told me, I don't cook the food for another person except for the child. (FGD #10, Respondent 7)*

These references to the caregivers' dialogue demonstrate how multiple avenues of instruction were important for knowledge translation and adherence to the recipe. For example, many caregivers, within all arms, recognized the connection of the strict instructions and their child's well-being.

> *They are not giving it to us to go and sell it, they are giving it to us for our children to develop. (FGD #14, Respondent 4)*

However, despite the willingness of caregivers to follow guidance discouraging sharing, this had drawbacks. For example, in the CSB+ w/oil and SC+A arms, not sharing within the household was described as creating potential conflicts, specifically that the food caused quarrels within the family. Others reported pressure to share in order to limit turmoil within the household.

> *This food has caused hate issues on me in the home. (FGD #4, Respondent 7)*

> *For this food that they give, my husband and I have dispute for it. (FGD #14, Respondent 8)*

> *There is somebody, when you collect this and you don't give them, they would not be happy. That is why some people do share it. She would share to her household members for them not to disagree with her. (FGD #6, Respondent 7)*

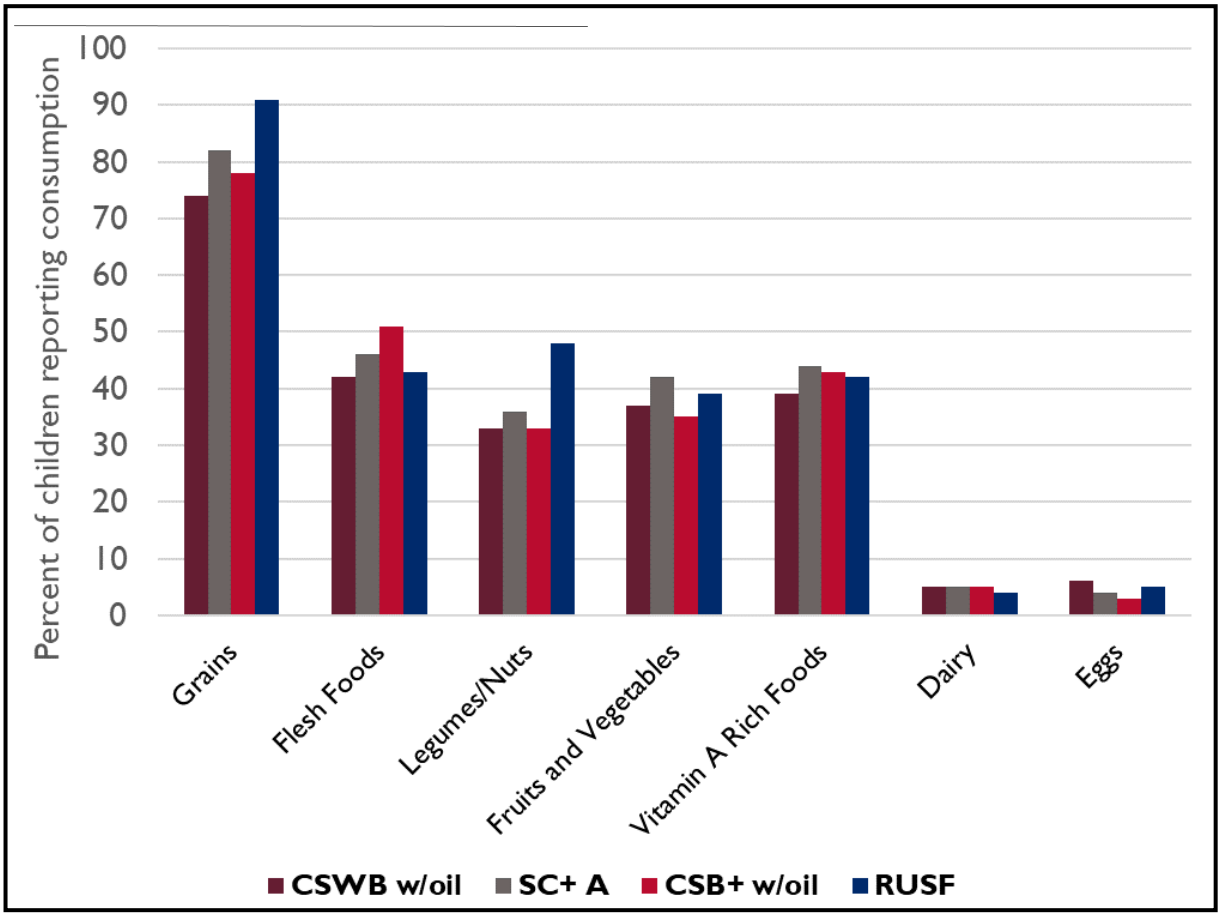
Similar sentiments were expressed in cases of sharing outside of the household in the CSB+ w/oil and SC+A arms. In two FGDs, caregivers expressed concerns about arguments among community members, especially in situations where social pressure was present.

> *If you tell someone that [you cannot share], you will have a quarrel with her. (FGD #1, Respondent 4)*

Displacement of Complementary Foods

There was minimal displacement of household complementary foods by the study foods in intention-to-treat analyses. In bivariate logistic regressions, children in the RUSF arm were significantly more likely to eat household grains (OR=3.55; 95% CI=2.02, 6.25) and legumes and nuts (OR=1.88; 95% CI=1.26, 2.80) than those in the other three arms. There were no significant differences in consumption of any of the other food groups by study arm. Reported food group consumption by study arm is shown in **Figure 22.**

Figure 22: Reported Food Group Consumption by Study Arm



General acceptability of the study foods among the caregivers and beneficiary children was illustrated by the lack of displacement of other foods, which was discussed across all arms in FGDs. Respondents discussed that the rationale for feeding the beneficiary child with other foods besides the study food was simply "doing as they were told." Other foods not mixed in with the study food but given in addition to it included rice, Bennimix[d] , cassava and mango. Breastfeeding was also seen as another form of feeding and was not omitted once the child was given the study food.

They said I should also serve my child with the food [CSB+] she normally eats, that is the only thing I add. (FGD #2, Respondent 2)

[d] Bennimix is a sesame-based flour manufactured in Sierra Leone: http://bennimixsl.com/

Displacement did occur minimally within the CSB+ w/oil and SC+A arms but was not evident in the other arms. In one SC+A FGD, when asked about the benefits of the study food, one caregiver responded with the benefit of not having to buy other food because the child can eat the study food.

> *Firstly, that food [the SC+A] has helped me, because I don't have to buy rice again, groundnut, benni, beans, and all those things. So I really take care of that food [the study food]. (FGD #8. Respondent 6)*

The main reason for giving the beneficiary child only the study food was because of shortages of other foods available in the house. One caregiver in the CSB+ w/oil arm expressed concern for not having any other available food in the house.

> *For me, if that happen [displacing other food], meaning there is no other food for my child, that is the time I will cook the pap [CSB+] again for my child because there is no other food. (FGD #15, Respondent 1)*

Food as Medicine

Across all study arms, caregivers perceived the study food as "medicine food." This was especially apparent among caregivers in the RUSF arm. Caregivers discussed that they believed the food would heal their child if they adhered to the instructions given. Many expressed their worry that mixing the food with other things would affect its medicinal qualities.

> *They [study staff] told us not to mix the food with other foods because you don't want the power of the drugs in the food to be destroyed or become less. (FGD #9, Respondent 1)*

Others used the foods' "medicinal qualities" as an excuse for not sharing the food with other children. Many caregivers believed that they were given medicine from the clinic, in the form of the study foods.

> *That one is a medicine food, so you should give it only to your child. (FGD #5, Respondent 3)*

> *It is already mixed with the ingredients…with medicines. The medicine that it was mixed with in the clinic. (FGD #4, Respondent 1)*

This perception was also visible throughout the preparation and cooking process, where many caregivers expressed their concern over the need to be very careful with the study food because of its medicinal qualities.

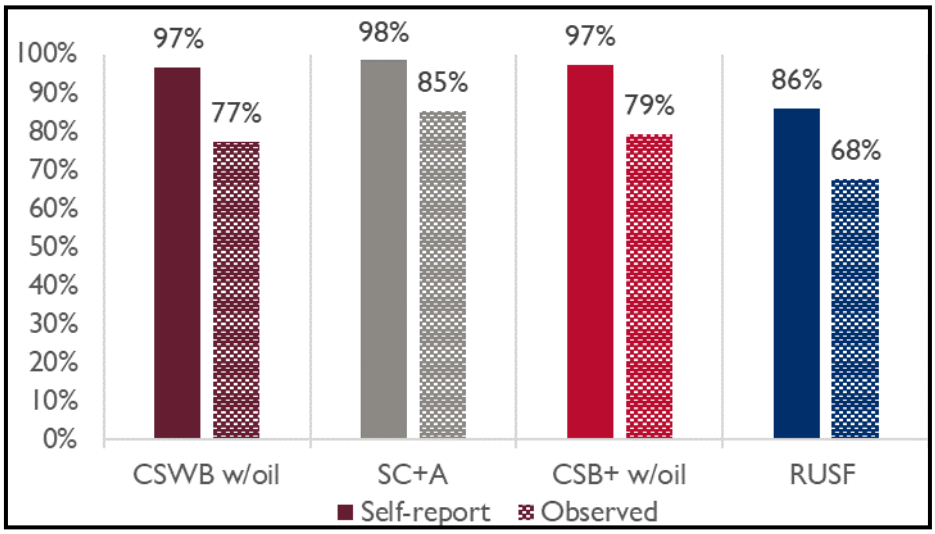
> *It should not cook too long so that the power will not die. (FGD #8, Respondent 6)*

The perception of the food as medicine is apparent throughout the caregivers' experience with the program. The perception of the food as medicine is displayed in **Appendix 4**.

b. Recipe Adherence Behavior

Figure 23 illustrates the differences between self-reported and observed behaviors in adhering to the recipe guidance. Self-reported was higher than observed adherence in all arms, varying from 13 to 20 percentage points between self-reported and observed behaviors. Though adherence to the recipe was overstated in the self-report compared to observed behavior, the general trends were consistent in both data collection methods. As expected, the CSB+ w/oil and CSWB w/oil arms performed similarly. The RUSF arm had the lowest rates of adhering to ingredient guidance, as a result of mixing the RUSF with rice porridge as a complementary food.

Figure 23: Percent of Caregivers Self-Reporting Adherence to Recipe Guidance Compared to Percent of Households Observed to Adhere at Least One Day in Five



Caregivers in the CSWB w/oil and CSB+ w/oil arms consistently reported using too much flour for an individual/daily ration. This contrasted with the SC+A study arm, where equal numbers of caregivers reported using too much, too little, and recommended quantities of flour (**Figure 24**). However, when comparing the ratio of water to flour, the pattern reversed (**Figure 25**): caregivers reportedly used too much flour in the SC+A arm, whereas caregivers receiving CSB+ w/oil and CSWB w/oil tended to use too much water. However, these variations did not influence recovery.

Figure 24: Percent of Caregivers Self-Reporting Using Correct Amount of Flour at Last Preparation Time

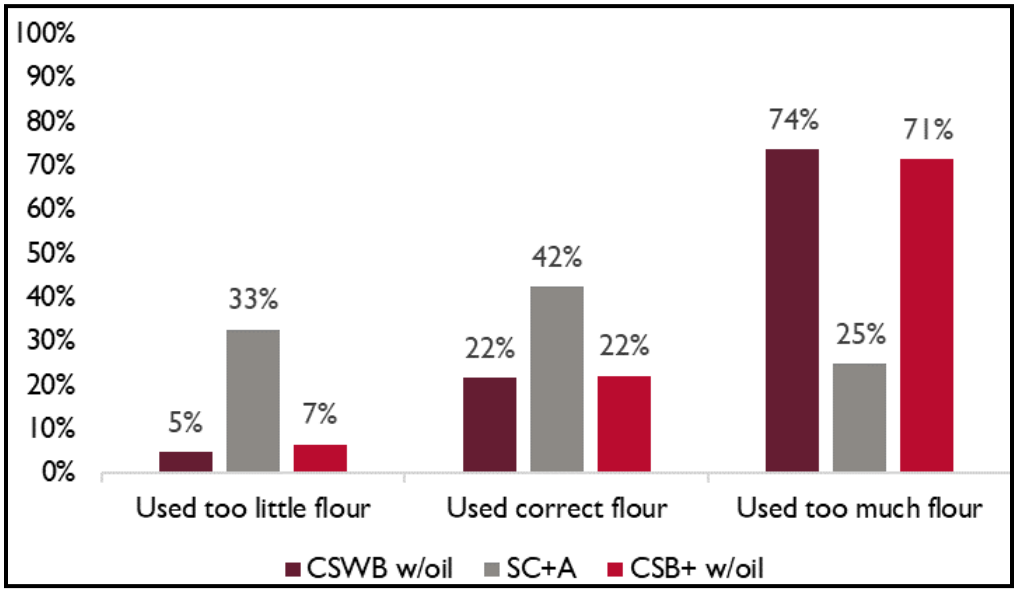
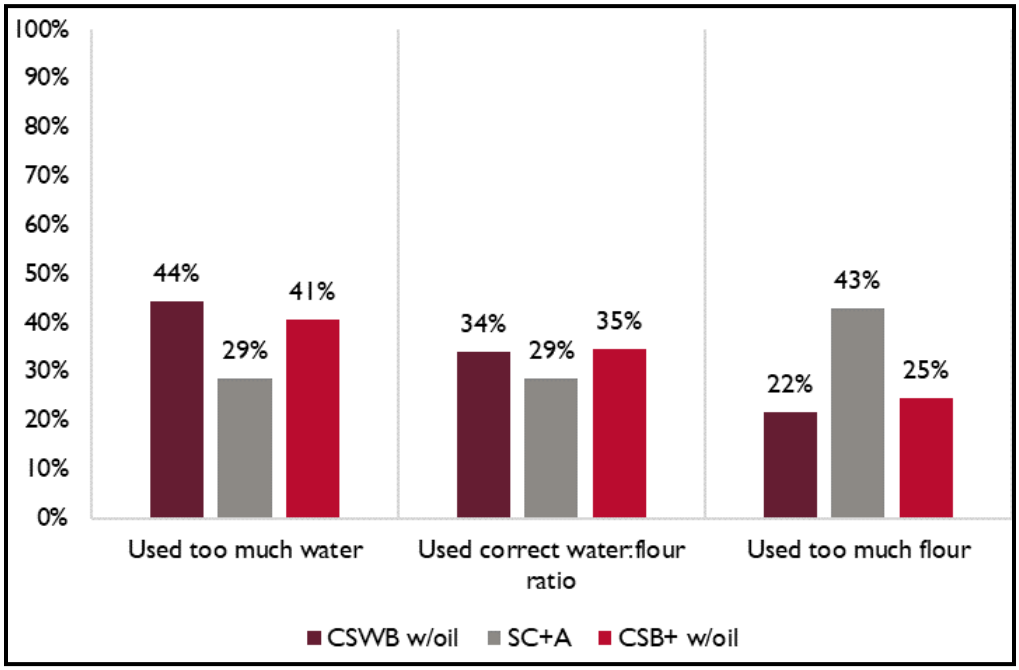


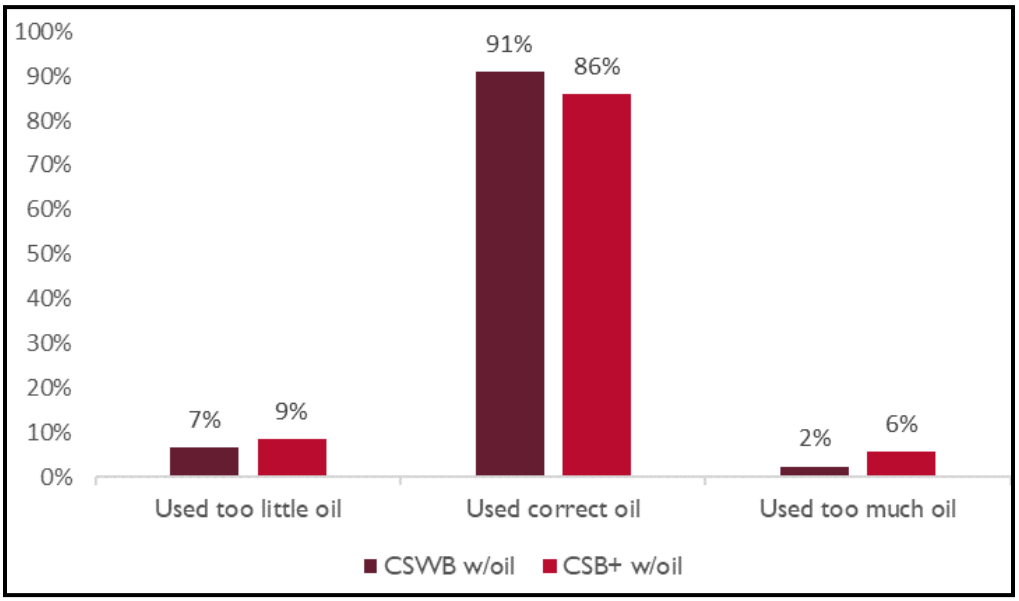
Figure 25: Percent of Caregivers Self-Reporting Using Correct Ratio of Flour: Water at Last Preparation Time



Among caregivers in the CSB+ w/oil and CSWB w/oil arms, nearly all caregivers reported using the correct amount of oil (**Figure 26**). Generally, as illustrated in **Figure 23**—in which more than 75% of households were observed to measure flour, oil, or water—caregivers did seem to be trying to adhere to the recipe. No differences were observed between older and younger children or between those breastfeeding and not breastfeeding, suggesting that differences in recipe adherence were dependent on

factors other than child's age. There were also no distinguishable differences in recipe adherence by outcome.

Figure 26: Percent of Caregivers Self-Reportedly Using Correct of Oil at Last Preparation Time



Knowledge Transfer

During FGDs, multiple themes emerged as either facilitators or barriers to recipe adherence. Main themes surrounding the caregivers' ability to adhere to the instructions included familial and community support, images of cooking steps on the packaging, and the recipe demonstrations. These themes helped facilitate knowledge translation of the instructions. Knowledge translation is most often defined in the literature as "the methods for closing the gaps from knowledge to practice".[45] For this analysis, knowledge translation included any references showcasing caregiver knowledge of how to cook the study food; verbalized actions that illustrated adherence to those instructions; or knowledge of other influencing practices such as intrahousehold sharing or food hygiene. Varying levels of familial and community support were also found to be an important facilitating factor. Familial support, both positive and negative, was defined as any household member that encouraged or hindered the caregiver's ability to adhere to and succeed in the program. Community support was also apparent. Community support was defined as support from any person outside the participant's immediate household that had a useful or obstructive effect on the caregiver's ability to succeed in the program.

Verbal instruction helped to reinforce the caregivers' ability to prepare the food as recommended. The use of photos to help with the caregivers' preparation process was most commonly expressed among the SC+A arm.

> *The packet that the blended [SC+A] is in, it has cups drawn on it, it is three and those cups directed us how to go about the food prepared, that is why they drawn on it, how we should go about it (preparing the food). (FGD #6, Respondent 2)*

> *If you do it [prepare the food] twice then you can boast of saying it so (that you are confident). If only you are a quick learner, just as the way the photos are explained, you can do it. (FGD #4, Respondent 1)*

Language or literacy as a barrier to understanding the photos was only apparent in the CSB+ w/oil arm. When asked about challenges related to the packaging, a few caregivers responded with individual strategies that could have helped them better understand the instructions, including utilizing positive community support as a resource. One went on to explain:

> *If you take the blended packet to someone who is educated and that person explains to you, you will understand. (FGD #7, Respondent 4)*

Husbands providing reminders to pick up the study food from the clinic was only reported in one SC+A FGD; the other FGDs discussed encouragement generally.

> *When I will be sleepy, my husband will shout at me saying, [caregiver name] don't you know you should go [pick up the food] today? (FGD #8, Respondent 2)*

Within each arm, husbands were seen as a resource for encouragement to follow the recipe. When asked about ways in which the caregivers felt supported, caregivers reported that husbands were useful for understanding the instructions and photos on the packaging because their literacy was higher.

> *My husband also did Mende and English learning. As soon as I reach home, then he removes the document [the packaging instructions] from my hand and sit on the veranda and read it for me. (FGD #8, Respondent 7)*

Caregivers in the FBF arms identified the recipe demonstration at enrollment to be the major facilitator for adhering to recipe guidance and understanding the preparation process. A common theme was the participants' ability to "boast" of their confidence in preparing the food the way they were instructed.

> *I will be confident to cook my child's pap at home just as how they taught me, that is how I am doing it at home. (FGD #6, Respondent 1)*

In the FBF arms, respondents offered explanations such as reasons why they felt confident in their ability to adhere to the demonstration instructions.

> *I have seen that my child is eating this pap [CSB+] after I cook it and she has improved. I am preparing it just as I was taught, she has been eating it and her body has changed. If that is the case, I can tell you that I know how to cook the food. (FGD #15, Respondent 5)*

In contrast, across all arms, there were barriers to adhering to instructions and food preparation shared by respondents. The most common theme identified was forgetfulness due to factors such as the duration of the demonstration and a general lack of attention during it. As expected, this was explicit in the FBF arms but not discussed in the RUSF arm. One caregiver, from the CSWB w/oil arm, explained that she could not remember the instructions because of the long duration of the demonstration.

> *I cannot really remember what they said because the explanation they gave us was very timely [long]. (FGD #12. Respondent 9)*

When asked about their initial visit to the clinic, one caregiver [CSB+ w/oil] discussed the length of the process, explaining that the long demonstration was not a good use of their time.

They took some time to teach us how to cook the blended before they gave it to us. This is what wasted our time. (FGD #7, Respondent 4)

10.3 Summary

Neither sharing, displacement, nor recipe adherence was associated with differences in recovery. Sharing was prevalent in all arms but highest among recipients of RUSF and lowest for SC+A. Similarly, differences between reported and observed sharing were highest in the RUSF arm. There were few differences in displacement among the four foods. Children who recovered were observed to consume the food most often compared to children who failed, who were observed to consume the food least often. Reported reasons for low rates of sharing included perceptions of the study food as medicine and of disease transference. Adherence to cooking instructions was high for flour and oil usage across FBFs. In SC+A, caregivers generally used less water or more flour than recommended. Reported facilitators to adherence were the recipe instructions and male/husband support in the home.

11. Challenges/Limitations

11.1 Logistical/Implementation Challenges

Lower than expected sample sizes meant that detectable differences were greater than originally anticipated. This occurred because of inaccurate data used at the scoping stage of the study indicating a higher MAM caseload than existed. This does not seem to have affected interpretation of the effectiveness analysis, in which the observed differences were small enough to suggest that they would not have been significant even if the planned sample size had been achieved.

Sierra Leone generally has a poor road infrastructure, and this was particularly true in Pujehun District. During the rainy season, roads would be underwater, and communities could be isolated from the rest of the district. This created challenges for the mobile SFP to reach clinics regularly during the rainy season and made data collection for the study challenging given the remote location of many caregivers. Though the teams worked diligently to reach and interview all sampled caregivers, it is possible that during the rainy season, the most remote potential respondents were not always reached.

Picture of enumeration team members crossing the Moa River during the dry season. Photo courtesy of Caritas Bo

11.2 Electronic Data Collection

Using tablets and electronic forms helped to decrease error in surveys. However, in a context such as Sierra Leone, where power is nearly nonexistent throughout the country, managing issues of device charging, data upload, and quirks of technology can leave enumerators in a situation where they are unable to collect data. These issues are easily managed with the provision of backup paper survey tools, but it can make training difficult when enumerators need to learn the same tool in multiple forms. In this study, there were only three documented cases of switching to paper forms (in one PHU observation and two IHOs).

11.3 Costing Limitations

This study was unable to collect detailed data on product losses. One objective of the smaller packaging was a reduction in losses of the FBFs, but there was no comparison to distribution from 25 kg bags often used in other settings. Absence of fumigation for the first tranche of foods at the main warehouse may have resulted in losses not captured by the study. Anecdotal evidence from caregivers showed that sometimes weevils were found in unopened bags of the FBFs. However, the frequency of these occurrences and their impact on cost-effectiveness remains unknown.

The novel packaging of CSB+ is an additional limitation of the cost-effectiveness analysis. Under traditional SFP modalities, 25 kg bags of CSB+ would have been repackaged in-country into smaller bags. This would have taken more of the staff and caregivers' time and could have resulted in spillage and spoilage losses of the FBFs. The impact on the cost-effectiveness analysis is unknown but should be considered when trying to generalize these findings to other SFP models in which the 25 kg bags would be used.

12. Summary of Findings

In this study set in southern Sierra Leone examining the relative effectiveness and cost-effectiveness of CSB+ w/oil, CSWB w/oil, SC+A, and RUSF, we found the four foods to be similarly effective and cost-effective when measured in cost per recovered child. Though caregivers' opportunity costs (caregiver perspective) were lower in the RUSF arm, the difference was not great enough to alter the comparative cost-effectiveness of the four foods when both caregiver and program perspective were combined. The mobile SFP's operational costs were the largest cost component of the SFP. This contrasts with other studies that found SNFs to be the largest cost component. [14] Unfortunately, as reported by Kennedy, et. al., are few cost-effectiveness studies of SFPs generally and even fewer that specifically report detail *on program costs*.[15]

Our findings are consistent with results from previous studies examining the effectiveness of different SNFs, in which neither RUFs nor FBFs performed systematically better than one another.[4,5] There are few studies exploring the comparative cost-effectiveness of different SNFs for MAM treatment, and they have reported different results.[6,7] Though previous studies cited lower caregiver opportunity costs as a rationale for increasing the use of RUFs in lieu of FBFs for MAM treatment, the findings of this study do not support these conclusions. [9,10]

With respect to the influence of the supply chain, we found that procuring commodities locally was not cheaper than international tenders. One novel element of this study was the use of smaller sized packaging for FBFs, which was more expensive but likely more efficient than the large paper-walled standard bags. Though there have been few formal studies examining the influence of different packaging types on the relative effectiveness of different SNFs, a recent review detailed the diverse challenges specific to SNF packaging.[16] In the review, the authors highlighted challenges with the 25 kg multiwall paper bags used in traditional settings and differences in the box sizes for the smaller, 1.25 kg bags

normal for SC+. Because this study did not use 25 kg bags of CSB+, we do not know if the smaller bags mitigated the challenges reported in the earlier review. However, the differences in box size was a challenge experienced at the storage warehouse when the study switched suppliers during the second tranche shipment.

Differences in effectiveness and cost-effectiveness emerged in the analysis of sustained recovery at one month, in which FBFs performed similarly to one another but differently from RUSF. This difference affected the cost-effectiveness point estimates of cost per child who sustained recovery—with RUSF achieving a lower rate of sustained recovery than FBFs, albeit not significantly—and this interpretation did not change when the caregiver perspective was also included. There are currently no published studies that have examined sustained recovery through a cost-effectiveness lens.

Regarding factors that may be associated with recovery, intrahousehold sharing of the foods was both reported and observed among all the foods. However, neither intrahousehold sharing nor following the recommended recipe affected the likelihood of recovery. *Consumption* was observed to be important; observed consumption was highest in households where children graduated. This study also found similar perceptions of the four foods as being medicinal and of high quality for improving the nutritional status of children. This positive association was cited as a rationale for sharing with children both within the home and within the broader community.

Previous studies argued that local perceptions of RUSF as a medicine led to a lower rate of sharing compared to FBFs.[28] In some studies, higher sharing rates (not reported) of FBFs was presumed to be a consequence of higher levels of acceptability among the beneficiary children receiving the FBF. [29] Programs often measure consumption by asking caregivers to return used sachets or bags.[46] In this study, we gauged consumption using retrospective self-reporting and five-day consecutive in-home observations, which may provide more accurate measures of intrahousehold feeding practices.

Finally, guidelines detailing porridge viscosity of FBFs derive from guidance on complementary feeding practices.[5] These guidelines, however, have not been evaluated with reference to recovery from acute malnutrition. The results reported here may indicated that adherence to feeding practices does not impact recovery from MAM. The relationship between sharing and recovery from acute malnutrition has also not been empirically explored in previous studies.[28] The findings of this study reveal that these factors may be of secondary importance to ensuring that the beneficiary consumes the supplementary food.

Acknowledgements

This study was made possible with the generous support of many partners, staff, and community members. In Sierra Leone, we want to thank the Community Action for the Welfare of Children (CAWeC) for their expertise and technical support in designing and implementing the social behavior change intervention for this study. Observational data on water, sanitation, and health practices in support of the environmental enteric dysfunction (EED) sub-study were made possible with the technical and logistical support of the Red Cross Society of Pujehun District, with our thanks and gratitude. The supplementary feeding program (SFP) and all data collection related to it were carried out by the staff of Project Peanut Butter (PPB). PPB also supported data collection for the EED, body composition, and neurocognitive sub-studies. Our thanks to Caritas Bo for managing and collecting all nonclinical data, including the survey and in-home observation data. Caritas Bo also accepted responsibility for data management of all main effectiveness study data.

The study was also supported technically with guidance, support, and supervision from the Pujehun District Nutrition Officer (Mrs. Mariama George) and the Pujehun District Medical Officers (Dr. David Bome and Dr. Teodros Gegbai). The success of the study was made possible with the support and experience of the 29 in-charge nurses and their catchment area community health workers. This District Health Management Team supported and made possible the study's integration with the Pujehun District health system and its supported communities. Our sincerest thanks also to the Pujehun District Council and the Paramount Chiefs of Pujehun District for their support and assistance with coordination and community entry.

Outside of Pujehun District, our sincerest thanks to Dr. Aiah Lebbie of Njala University for the use of liquid nitrogen and to the Infectious Disease Research Laboratory at the University of Makeni and its highly qualified staff for support in storing our biological samples. Preliminary analysis of the porridge samples was conducted by the Sierra Leone Standards Bureau. Our sincerest thanks to WFP/Sierra Leone for their generous logistical support in managing the four study foods both in Kenema and in Freetown. To our partners at the Directorate of Food and Nutrition in the Ministry of Health and Sanitation and Ms. Aminata S. Koroma, thank you for your support of the study and the use of the results to further the fight against malnutrition in Sierra Leone.

In the United States, our thanks to Dr. Bill Wong of Baylor College of Medicine and his team's assistance with analyzing and interpreting the deuterium dilution samples for the body composition sub-study. Our thanks also to Dr. Sridevi Devaraj at Baylor College of Medicine for analyzing the lactulose to mannitol samples for the EED sub-study. Thanks to colleagues at Washington University in St. Louis for support in analyzing and interpreting the mRNA samples, and similarly to the Nutritional Immunology Laboratory and Human Nutrition Research Center on Aging at Tufts University for the fecal protein sample analysis, and to Dr. Anne Kane at the Tufts Medical Center for analysis of the 16 mRNA samples for microbiota, all for the EED sub-study. Finally, thanks to Judy Canahuati and Rufino Perez of USAID for their support in the design, implementation, and dissemination of the Sierra Leone Four Foods Treatment Study.

References

1. Webb P, Lorge Rogers B, Rosenberg I, et al. *Improving the Nutritional Quality of U.S. Food Aid: Recommendations for Changes to Products and Programs*.; 2011. https://pdf.usaid.gov/pdf_docs/pnadz841.pdf. Accessed August 14, 2018.
2. Radhakrishna KV, Kulkarni B, Balakrishna N, Rajkumar H, Omkar C, Shatrugna V. Composition of weight gain during nutrition rehabilitation of severely under nourished children in a hospital based study from India. *Asia Pac J Clin Nutr*. 2010;19(1):8-13. doi:10.6133/apjcn.2010.19.1.02
3. Campbell RK, Schulze KJ, Shaikh S, et al. Environmental enteric dysfunction and systemic inflammation predict reduced weight but not length gain in rural Bangladeshi children. *Br J Nutr*. 2018;119(4):407-414. doi:10.1017/S0007114517003683
4. Annan RA, Webb P, Brown R. Management of Moderate Acute Malnutrition (MAM): Current Knowledge and Practice Collaborating to improve the management of acute malnutrition worldwide. 2014;(September). www.cmamforum.org.
5. Lazzerini M, Rubert L, Pani P. Specially formulated foods for treating children with moderate acute malnutrition in low- and middle-income countries (Review). 2013;(6). doi:10.1002/14651858.CD009584.pub2.www.cochranelibrary.com
6. Isanaka S, Barnhart DA, McDonald CM, et al. Cost-effectiveness of community-based screening and treatment of moderate acute malnutrition in Mali. *BMJ Glob Heal*. 2019;4(2):1-10. doi:10.1136/bmjgh-2018-001227
7. Suri DJ, Moorthy D, Rosenberg IH. The Role of Dairy in Effectiveness and Cost of Treatment of Children with Moderate Acute Malnutrition. *Food Nutr Bull*. 2016;37(2):176-185. doi:10.1177/0379572116633327
8. Puett C, Sadler K, Alderman H, Coates J, Fiedler JL, Myatt M. Cost-effectiveness of the community-based management of severe acute malnutrition by community health workers in southern Bangladesh. *Health Policy Plan*. 2013;28(4):386-399. doi:10.1093/heapol/czs070
9. Nackers F, Broillet F, Oumarou D, et al. Effectiveness of ready-to-use therapeutic food compared to a corn/soy-blend-based pre-mix for the treatment of childhood moderate acute malnutrition in Niger. *J Trop Pediatr*. 2010;56(6):407-413. doi:10.1093/tropej/fmq019
10. Lagrone L, Cole S, Schondelmeyer A, Maleta K, Manary MJ. Locally produced ready-to-use supplementary food is an effective treatment of moderate acute malnutrition in an operational setting. *Ann Trop Paediatr*. 2010;30(2):103-108. doi:10.1179/146532810X12703901870651
11. WHO. WHO, UNICEF, WFP and UNHCR Consultation on the Programmatic Aspects of the Management of Moderate Acute Malnutrition in Children under five years of age. 2010;(February):1-18.
12. Fabiansen C, Yameogo CW, Iuel-Brockdorf AS, et al. Effectiveness of food supplements in increasing fat-free tissue accretion in children with moderate acute malnutrition: A randomised 2 x 2 x 3 factorial trial in Burkina Faso. *PLoS Med*. 2017;14(9):e1002387. doi:10.1371/journal.pmed.1002387
13. Mcdonald CM, Ackatia-armah RS, Doumbia S, Kupka R, Duggan CP, Brown KH. Percent Fat Mass Increases with Recovery , But Does Not Vary According to Dietary Therapy in Young Malian Children Treated for Moderate Acute Malnutrition. 2019.

14. Isanaka S, Barnhart DA, McDonald CM, et al. Cost-effectiveness of community-based screening and treatment of moderate acute malnutrition in Mali. *BMJ Glob Heal*. 2019;4(2):e001227. doi:10.1136/bmjgh-2018-001227

15. Kennedy E, Branca F, Webb P, Bhutta Z, Brown R. Setting the scene: An overview of issues related to policies and programs for moderate and severe acute malnutrition. *Food Nutr Bull*. 2015;36(1):S9-S14. doi:10.1177/15648265150361S102

16. Roubert A, Johnson Q, Walton S, Webb P. *FOOD AID PACKAGING CHALLENGES AND OPPORTUNITIES : A REVIEW OF THE PACKAGING OF FORTIFIED VEGETABLE OIL , CORN SOY BLEND PLUS , AND SUPER CEREAL PLUS A Report from the Food Aid Quality Review*. Boston, MA; 2018.

17. Heirman BJ, Jenkins M, Rosenzweig J. Lessons learned from evaluations of the impact of WFP programmes on moderate acute malnutrition in the Sahel. 2019;(60):107-110.

18. Osendarp S, Rogers B, Ryan K, et al. Ready-to-use foods for management of moderate acute malnutrition: Considerations for scaling up production and use in programs. *Food Nutr Bull*. 2015;36(1):S59-S64. doi:10.1177/15648265150361S110

19. Wegner CW, Loechl C, Mokhtar N. Moderate acute malnutrition: Uncovering the known and unknown for more effective prevention and treatment. *Food Nutr Bull*. 2015;36:S3-S8. doi:10.1177/15648265150361S101

20. Pickering AJ, Null C, Winch PJ, et al. The WASH Benefits and SHINE trials: interpretation of WASH intervention effects on linear growth and diarrhoea. *Lancet Glob Heal*. 2019;7(8):e1139-e1146. doi:10.1016/S2214-109X(19)30268-2

21. Humphrey J, N Mbuya M, Moulton L, et al. Independent and combined effects of improved water, sanitation, and hygiene, and improved complementary feeding, on child stunting and anaemia in rural Zimbabwe: a cluster-randomised trial. *Lancet Glob Heal*. 2019;7:132-179. doi:10.1016/S2214-109X(18)30374-7

22. Friis H, Michaelsen KF, Wells JC. Choice of design and outcomes in trials among children with moderate acute malnutrition. *Food Nutr Bull*. 2015;36(1):S35-S40. doi:10.1177/15648265150361S106

23. Chang CY, Trehan I, Wang RJ, et al. Children Successfully Treated for Moderate Acute Malnutrition Remain at Risk for Malnutrition and Death in the Subsequent Year after Recovery. *J Nutr*. 2013;143(2):215-220. doi:10.3945/jn.112.168047

24. Trehan I, Banerjee S, Murray E, et al. Extending Supplementary Feeding for Children Younger Than 5 Years With Moderate Acute Malnutrition Leads to Lower Relapse Rates. *J Pediatr Gastroenterol Nutr*. 2015;60(4):544-549. doi:10.1097/MPG.0000000000000639

25. Stobaugh HC, Bollinger LB, Adams SE, et al. Effect of a package of health and nutrition services on sustained recovery in children after moderate acute malnutrition and factors related to sustaining recovery: A cluster-randomized trial. *Am J Clin Nutr*. 2017;106(2):657-666. doi:10.3945/ajcn116.149799

26. Black RE, Allen LH, Bhutta ZA, et al. Maternal and child undernutrition: global and regional exposures and health consequences. *Lancet*. 2008;371(9608):243-260. doi:10.1016/S0140-6736(07)61690-0

27. Ruel MT, Alderman H. Nutrition-sensitive interventions and programmes: How can they help to accelerate progress in improving maternal and child nutrition? *Lancet*. 2013;382(9891):536-551. doi:10.1016/S0140-6736(13)60843-0

28. Karakochuk CD, van den Briel T, Stephens D, Zlotkin S. Food Sharing Practices in Households Receiving Supplemental Foods for the Treatment of Moderate Acute Malnutrition in Ethiopian Children. *J Hunger Environ Nutr*. 2015;10(3):343-355. doi:10.1080/19320248.2014.962772

29. Matilsky DK, Maleta K, Castleman T, Manary MJ. Supplementary Feeding with Fortified Spreads Results in Higher Recovery Rates Than with a Corn/Soy Blend in Moderately Wasted Children. *J Nutr*. 2009;139(4):773-778. doi:10.3945/jn.108.104018

30. Casey AR. *U . S . International Food Assistance: An Overview*. Washington, D.C.; 2018. https://fas.org/sgp/crs/row/R45422.pdf.

31. USAID. Office of Food For Peace *Fiscal Year 2018 Annual Report*. Washington, D.C.; 2018. https://www.usaid.gov/sites/default/files/documents/1866/FY2018_FFP_Annual_Report_508_compliant.pdf.

32. Rogers BL, Wilner LB, Maganga G, et al. Program changes are effective and cost-effective in increasing the amount of oil used in preparing corn soy blend porridge for treatment of moderate acute malnutrition in Malawi. *Matern Child Nutr*. 2017;13(4):1-10. doi:10.1111/mcn.12393

33. Cliffer I, Suri D, Langlois B, et al. *Comparative Cost-Effectiveness of Four Supplementary Foods in Preventing Stunting and Wasting in Children 6-24 Months in Burkina Faso. Report to USAID*. Boston, MA; 2018.

34. Lenters LM, Wazny K, Webb P, Ahmed T, Bhutta ZA. Treatment of severe and moderate acute malnutrition in low-and middle-income settings: a systematic review, meta-analysis and Delphi process. *BMC Public Health*. 2013;13 Suppl 3(Suppl 3):S23. doi:10.1186/1471-2458-13-S3-S23

35. Webb P, Rogers B, Rosenberg I, et al. *Delivering Improved Nutrition: Recommendations for Changes to U.S. Food Aid Products and Programs*. Vol 37.; 2011. doi:10.1177/0379572116629780

36. Broersen B, Sinamo S, Nsabimana J, et al. Amylase increases energy and nutrient density of Super Cereal Plus porridge as prepared and accepted by Rwandan caregivers. *Matern Child Nutr*. 2018;(October 2018). doi:10.1111/mcn.12742

37. Kampstra NA, Van Hoan N, Koenders DJPC, et al. Energy and nutrient intake increased by 47–67% when amylase was added to fortified blended foods—a study among 12- to 35-month-old Burkinabe children. *Matern Child Nutr*. 2018;14(1):1-12. doi:10.1111/mcn.12459

38. USAID, UCLA School of Public Health Nutrition Assessment Unit. Sierra Leone : national nutrition survey : final report. 2017.

39. Statistics Sierra Leone (SSL) and ICF International. *Sierra Leone Demographic and Health Survey 2013*. Freetown, Sierra Leone and Rockville, Maryland, USA:; 2014. doi:10.4324/9780203403099_Sierra_Leone

40. Statistics Sierra Leone. Sierra Leone Multiple Indicator Cluster Survey 2017: Survey findings report. 2017.

41. Quee, Daniel; Tucker, Harrison; Koroma, Mariatu; Griswold, Stacy; Walton, Shelley; Suri, Devika;

Langlois, Breanne; Rogers B. *Cooking Instruction Development and Acceptability Tests of Corn-Soy Blend Porridges : Pujehun District , Sierra Leone*. Boston, MA; 2016.

42. Iuel-Brockdorf AS, Dræbel TA, Fabiansen C, et al. Acceptability of new formulations of corn-soy blends and lipid-based nutrient supplements in Province du Passoré, Burkina Faso. *Appetite*. 2015;91:278-286. doi:10.1016/j.appet.2015.04.058

43. Glaser, B. G., & Strauss AL. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. New Brunswick: Aldine: Routledge; 2009.

44. Baltussen, Rob M. P. M; Adam, Taghreed; Tan-Torres Edejer, Tessa; Hutubessy, Raymond C. W; Acharya, Arnab; Evans, David B; Murray CJL. *Making Choices in Health : WHO Guide to Cost-Effectiveness Analysis*. (Tan-Torres Edejer T. et. al., ed.). Geneva: Springer Netherlands; 2003. https://apps.who.int/iris/bitstream/handle/10665/42699/9241546018.pdf?sequence=1&isAllowed=y .

45. Straus SE, Tetroe J, Graham I. Defining knowledge translation. *Cmaj*. 2009;181(3-4):165-168. doi:10.1503/cmaj.081229

46. Style S, Tondeur M, Wilkinson C, et al. Operational Guidance on the use of special nutritional products in refugee populations. *Food Nutr Bull*. 2013;34(4):420-428. doi:10.1177/156482651303400407

Appendix 1: Four Foods Formulations

Nutrients	Unit	Super Cereal Plus (SC+) (Comparison)	Corn Soy Blend Plus (CSB+) with Fortified Vegetable Oil (FVO)		Corn Soy Whey Blend (CSWB) and Fortified Vegetable Oil (FVO)		Ready-to-Use Supplementary Food (RUSF)	
Serving (grams)		135.7 g/day	85.7 g CSB+, 25.7 g FVO/day	85.7 g CSB+, 25.7 g FVO/day	85.7 g CSWB, 25.7 g FVO/day	85.7 g CSWB, 25.7 g FVO/day	RUSF 100 g/day	RUSF 100 g/day
Repackaged Bag Size (2 weeks)		1.9 kg bag SC+	1.2 kg bag CSB+ and 0.36 kg bottle of oil	1.2 kg bag CSB+ and 0.36 kg bottle of oil	1.2 kg bag CSWB and 0.36 kg bottle of oil	1.2 kg bag CSWB and 0.36 kg bottle of oil	N/A	N/A
			Min	**Max**	**Min**	**Max**	**Min**	**Max**
Energy Minimum	kcal	556.37	552.85	552.85	552.85	552.85	510.00	560
Protein	g	21.71	12.00	12.00	12.00	12.00	11.00	16
Fat	g	12.21	30.84	30.84	30.84	30.84	26.00	36
Vitamin A	IU	4,703.36	4,507.22	4,892.72	4,509.39	4,894.89	1,833.33	3833.33
Niacin	mg	10.86	6.86	6.86	6.86	6.86	13.00	_
Pantothenic Acid	mg	2.17	1.37	1.37	1.37	1.37	4.00	_
Vitamin B6	mg	1.36	0.86	0.86	0.86	0.86	1.80	1.80
Folate	mcg	149.27	94.27	94.27	94.27	94.27	_	_
Vitamin B12	mcg	2.71	1.71	1.71	1.71	1.71	2.70	2.70
Vitamin C	mg	122.13	77.13	77.13	77.13	77.13	60.00	60
Vitamin D	mcg		411.20	591.10	411.20	591.10	15.00	20
Vitamin D3	IU	599.25	378.45	378.45	378.45	378.45	_	_
Vitamin E	mg	11.26	9.22	9.22	9.22	9.22	16.00	_
Vitamin K	mcg	40.71	72.97	72.97	72.97	72.97	27.00	_
Vitamin B1 (Thiamine)	mg	0.27	0.17	0.17	0.17	0.17	1.00	_
Vitamin B2 (Riboflavin)	mg	1.90	1.20	1.20	1.20	1.20	2.10	_
Iron (Ferrous fumarate)	mg	5.43	3.43	3.43	3.43	3.43	10.00	14
Iron (Iron-sodium EDTA)	mg	3.39	2.14	2.14	2.14	2.14	_	_

			Min	Max	Min	Max	Min	Max
Zinc	mg	6.79	4.29	4.29	4.29	4.29	11.00	14
Iodine	mcg	54.28	34.28	34.28	34.28	34.28	100.00	140
Potassium	mg	189.98	119.98	119.98	119.98	119.98	900.00	1400
Phosphorus	mg	314.82	248.53	248.53	239.96	239.96	450.00	750
Calcium	mg	613.36	387.36	387.36	310.23	310.23	535.00	750
Biotin	mcg	_	7.03	7.03	_	_	60.00	_
Copper	mg	_	_	_	_	_	1.40	1.9
Magnesium	mg	_	_	_	_	_	150.00	225
Manganese	mg	_	_	_	_	_	1.20	_
Selenium	mcg	_	_	_	_	_	20.00	40
Sodium	mg	_	_	_	_	_	_	_

Appendix 2: Statistical Methods

PHU Sampling Process

The sampling frame consisted of 72% in the accessibility stratum and 27% in the inaccessibility stratum. To remain consistent with this distribution of accessible/inaccessible PHUs, the first step of the sampling process randomly selected 20 PHUs from the accessible stratum and 8 from the inaccessible stratum from the sampling frame. Next, 8 (or 20) block IDs were randomly assigned to the PHUs four copies of the sequence 1 through 2 for inaccessible, and four copies of sequence 3 through 7 for accessible, representing the individual blocks among which randomization was carried out. From the resultant sampling plan, statistics were generated using the stratification criteria above (e.g., total number of chiefdoms, total number of types of PHUs included in the plan, coefficients of variation in total MAM cases within each of the seven blocks, number of PHUs with mines).

Hundreds of thousands of potential plans were simulated through a computer program and reduced as more criteria were applied. One simulation was randomly selected from those that met all the criteria. Then, another round of sampling was carried out. For each block of 4, a group indicator was randomly assigned (ranging from 1 to 4, representing each of the four treatments). From this, the total number of types of PHUs in each group and total sum of MAM cases in each group were calculated. This process was repeated thousands of times using the same 28 selected PHUs and then pared down using the following predetermined criteria: every group had to contain all three types of PHUs, and the coefficient of variation for the MAM cases across the four groups needed to be lower than 0.1. This generated a set of 351 possible randomization schemes. From this set of 351, one was randomly selected as the sampling scheme for the study.

Further scoping was done by the field research manager to see if there were any obstacles to inclusion of each clinic. It was found that several PHUs offered outpatient therapeutic programs (OTPs) for SAM, which was determined to be of importance. Therefore, the 28 PHUs were reshuffled into the four groups in order to have a more even allocation of PHUs that offered OTPs across arms. Some clinics had to be exchanged because they were located too close to other PHUs or were only accessible by boat. Clinics were exchanged one by one by going back to the full PHU list and selecting replacements in order of descending MAM cases until no further issues were noted and the seven clinics in each study arm appeared to be balanced based on the selection criteria. Later in the study, due to low enrollment rates, expanded catchment schemes were put into place. In 2018, an eighth PHU was added to the CSWB arm to improve balance between the arms. See **Figure 3** for a map of Pujehun District with the PHU locations.

Statistical Modeling

Mixed-effect logistic regression models were fit using PHU as the random effect. The binary outcome variable was defined as achieving MUAC ≥ 12.5cm within 12 weeks of rations received versus not. Crude and adjusted models were fit. Observations with missing covariate data were dropped from the models. To evaluate whether missing data had an impact on parameter estimates, an additional model with multiple imputations was fit to predict the missing values from the rest of the data. The adjusted models with and without multiple imputations were compared. After the regression models were estimated, the intraclass correlation was calculated, and marginal predictions of the study arm variable were obtained. Crude and adjusted models were also fit to exclude those who missed three consecutive visits (i.e., defaulted) in order to evaluate whether the study foods differed among the sample who adhered to the treatment program.

Multicollinearity of model covariates was assessed using a variance inflation factor: if a covariate had a value above 15, it was removed from the model. The link test was used to determine if there was specification error. To evaluate for presence of influential observations and outliers that may have affected parameter estimates, plots of Pearson and deviance residuals and the Pregibon delta beta influence were examined.

For the Cox Hazards Models, prior to modeling, univariate analyses were conducted to check the Kaplan-Meier curves for each potential categorical predictor, checking the shapes of the survival functions and assessing proportionality. Tests of equality across strata (levels of each variable) were also conducted to identify potential variable candidates for the final model, using univariate Cox proportional hazards regressions for each continuous predictor and nonparametric log-rank test of equality to identify categorical predictor candidates. Variables for which the p-value was 0.05 or less were included in the final model. Interactions of each predictor and the time variable (days to recovery) were used to test violations of the proportionality assumption (i.e., that the ratio of hazards for any two individuals is constant over time), and final models were checked for goodness-of-fit using Cox-Snell residuals.

Appendix 3: Cost from the Program Perspective (Additional Scenarios)

As explained in **Table 15**, the research team used additional data sources to explore how costs from the program perspective compared across arms in three product specification and procurement channel scenarios. This was in addition to scenario 1, which served as the basis for cost-effectiveness analysis.

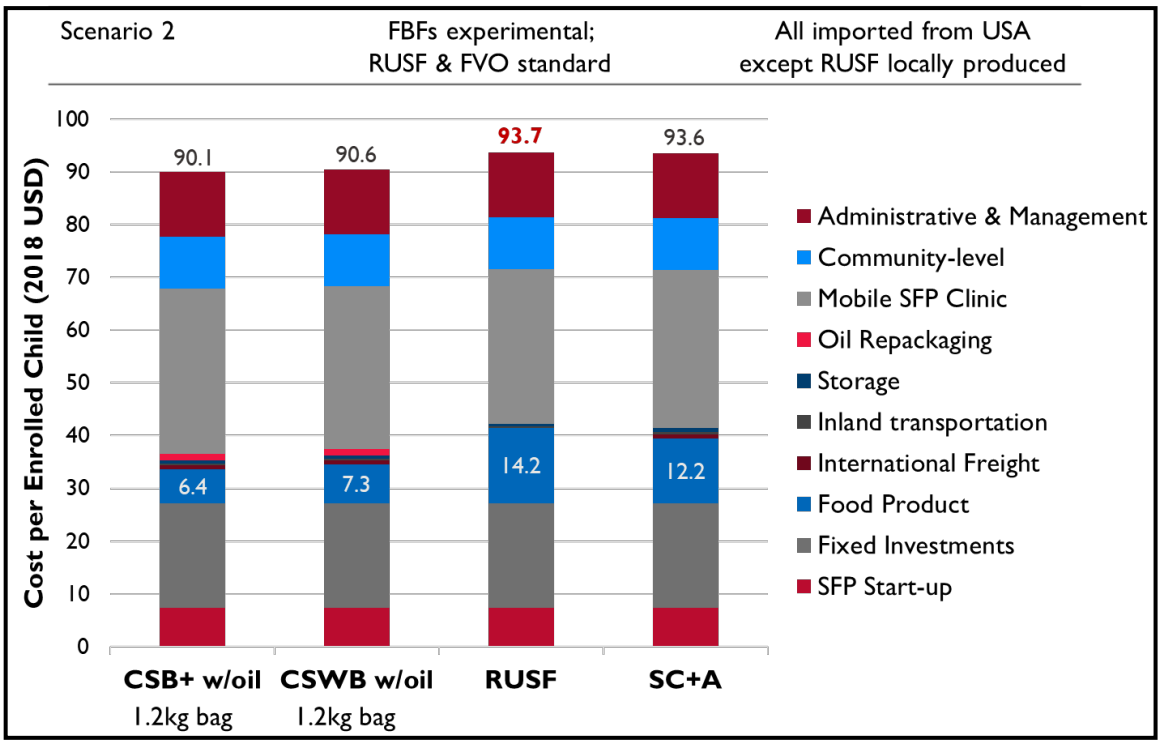
Table 15: Product Specification and Procurement Scenarios

	Product Specification: Experimental versus Standard	Procurement Channel: Imported versus LRP
Scenario 1 (Primary)	FBFs experimental; RUSF & FVO standard ı	All imported from USA
Scenario 2	FBFs experimental; RUSF & FVO standard ı	All imported from USA except RUSF locally procured ı
Scenario 3	SC+A and CSWB experimental formulation only; All standard packaging	All imported from USA
Scenario 4	All standard (exclude CSWB in this scenario)	All locally procured

Acronyms: LRP: Local and Regional Procurement; FBF: Fortified Blended Flours, CSWB: Corn Soy Whey Blend; SC+A: Super Cereal Plus with amylase; RUSF: Ready-to-Use Supplementary Food; FVO: Fortified Vegetable Oil

ı the scenario has the same product specification or channel as incurred in the study

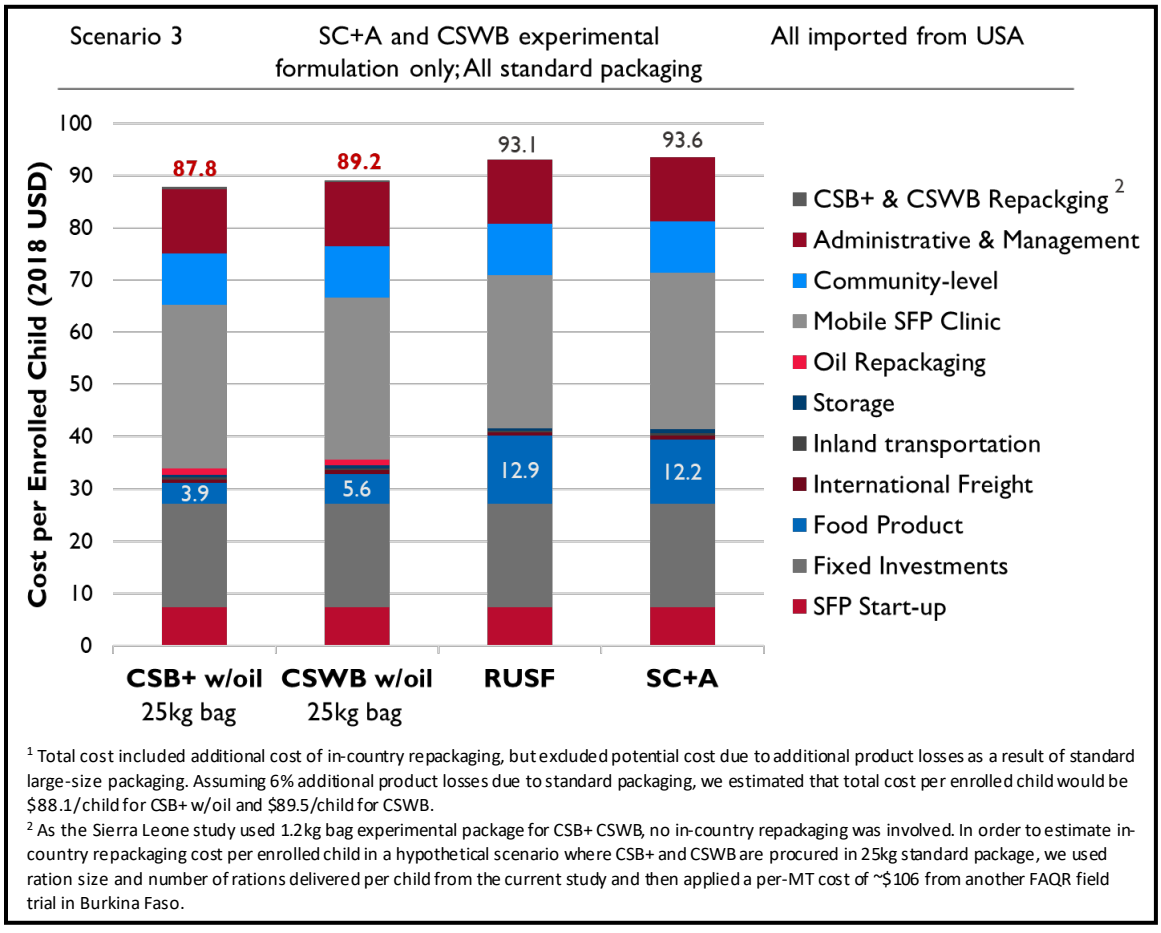
Figure 27: Program Perspective Cost per Enrolled Child (Scenario 2)



Scenario 2 compared the four arms using the actual specifications and procurement channels used for the study. For the locally produced RUSF, international freight costs dropped to zero while product costs increased (**Table 11** and **Table 12**). **Figure 27** shows that, compared to the primary scenario, the net difference in total cost per enrolled child for RUSF was small, amounting to about $0.60 per child increase.

Scenario 3 explored the cost impact of packaging specification changes for CSB+ and CSWB. Compared with the primary scenario using experimental 1.2 kg bags, program perspective costs per enrolled child using standard 25 kg bags decreased by about $2 per child for the two arms (**Figure 28**). This cost calculation for scenario 3 included additional in-country repackaging using cost estimates from FAQR Burkina Faso trial.[33]

Figure 28: Program Perspective Cost per Enrolled Child (Scenario 3) [1]



[1] Total cost included additional cost of in-country repackaging, but excluded potential cost due to additional product losses as a result of standard large-size packaging. Assuming 6% additional product losses due to standard packaging, we estimated that total cost per enrolled child would be $88.1/child for CSB+ w/oil and $89.5/child for CSWB.

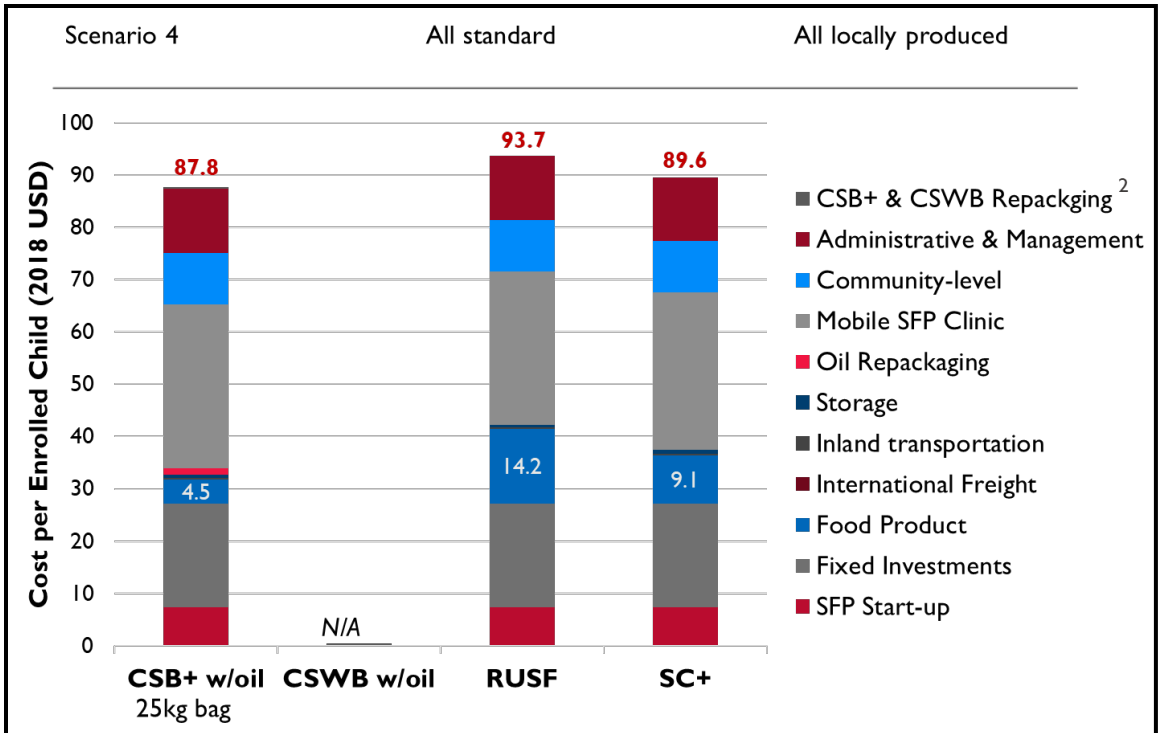
[2] As the Sierra Leone study used 1.2kg bag experimental package for CSB+ CSWB, no in-country repackaging was involved. In order to estimate in-country repackaging cost per enrolled child in a hypothetical scenario where CSB+ and CSWB are procured in 25kg standard package, we used ration size and number of rations delivered per child from the current study and then applied a per-MT cost of ~$106 from another FAQR field trial in Burkina Faso.

Additional product losses could occur due to spillage in distributions/repackaging and spoilage in storage when CSB+ and CSWB are in standard paper 25 kg packaging. To assess possible impact of food losses on cost, the research team assumed 6% losses throughout the supply chain and estimated an additional slight increase in total cost per enrolled child of ~$0.30 per child for the CSB+ w/oil and CSWB w/oil arms. We used 6% based on a previous FAQR trial in Burkina Faso that evaluated CSB+ and CSWB in standard packaging for an 18-month blanket supplementary feeding program.[33]

A total of ~2.3% losses for CSB+ and ~6.2% losses for CSWB were observed during multiple supply chain steps (we did not have data on losses during distribution). However, due to differences in supply chain conditions, local contexts, and the duration of stay for the foods between the two trials, it would be difficult to predict the exact amount of additional product losses if standard packaging had been used. Therefore, the research team calculated the turning points where total costs per enrolled child for CSB+ w/ oil and CSWB w/ oil using standard packaging would no longer be cheaper than using experimental packaging. Those turning points for percent losses were 43% for CSB+ and 25% for CSWB. As product and supply chain were not the largest components of total cost in this SFP for MAM treatment, the influence of additional losses on cost per enrolled child due to packaging was relatively small for CSB+ w/oil and CSWB w/oil.

In Scenario 4, CSB+ w/oil, RUSF, and SC+ were compared in standard specifications to explore the cost implications of local procurement (**Figure 29**). For CSB+ w/oil, total cost per enrolled child remained the same as in scenario 3, because the decrease in international freight was offset by the increase in product cost. For SC+, the total cost per enrolled child decreased by $4 per child compared to scenario 3, in part due to the lack of amylase in the LRP version. Therefore, in this scenario, total cost per enrolled child rankings changed substantially: SC+ became more comparable to CSB+ w/oil, and RUSF became the most expensive option. The price estimates for CSB+ and SC+ procured by WFP came from African suppliers outside of West Africa. It is unclear whether local capacity exists yet to produce FBFs or if they would be sold at similar price levels.

Figure 29: Program Perspective Cost per Enrolled Child (Scenario 4)[1]



[1] Total cost included additional cost of in-country repackaging, but excluded potential cost due to additional product losses as a result of standard large-size packaging. Some of the LRP prices (SC+ and CSB+) were derived from suppliers outside of West Africa, but we assumed West Africa suppliers would be able to produce these products at similar price levels. Otherwise, additional freight cost would apply.

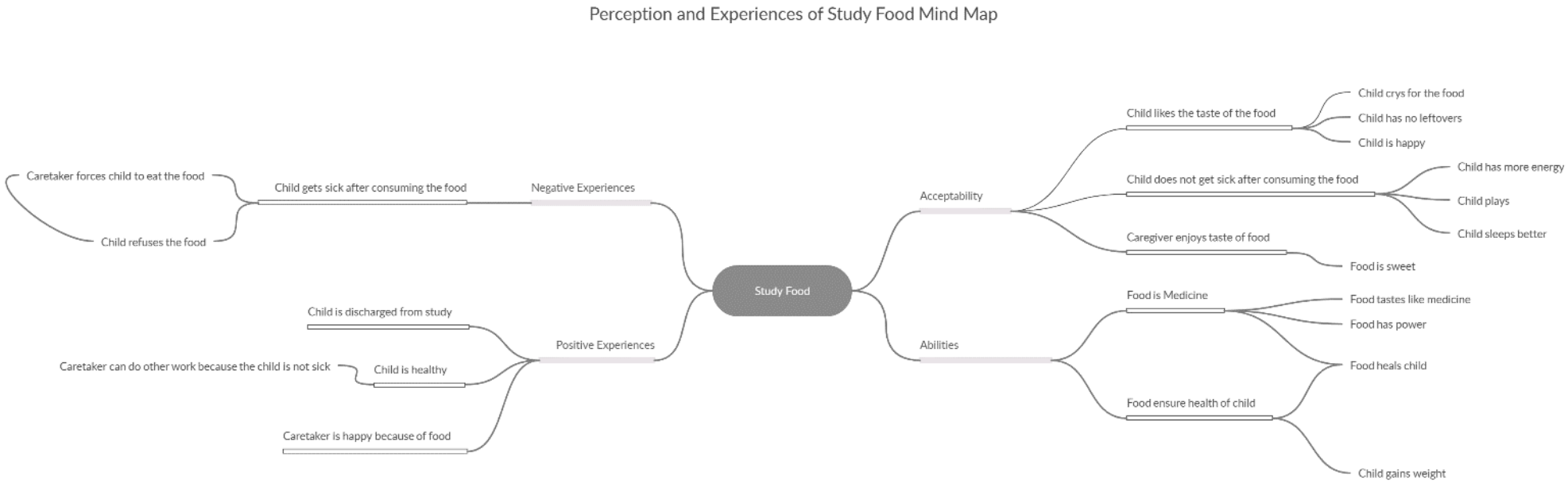
[2] As the Sierra Leone study used 1.2kg bag experimental package for CSB+ CSWB, no in-country repackaging was involved. In order to estimate in-country repackaging cost per enrolled child in a hypothetical scenario where CSB+ and CSWB are procured in 25kg standard package, we used ration size and number of rations delivered per child from the current study and then applied a per-MT cost of ~$106 from another FAQR field trial in Burkina Faso.

Summary

If CSB+ were packaged in standard 25 kg bag, this would have increased the cost difference (i.e., reduced the cost per enrolled child of CSB+ w/oil) compared to RUSF and SC+A. However, easier in-country handling for both implementers and caregivers as well as potential prevention of product losses should also be factored into packaging decisions for CSB+. Potential ways to reduce the cost of small packaging for CSB+ should be explored.

Product procurement channel/location did not make a substantial difference in total cost comparisons. This is because the elimination of international freight cost was mostly offset by higher in-country inspection cost of locally produced SNFs.

Appendix 4: Perceptions of Study Food (Qualitative Results)



Appendix 5: Reasons for Sharing and Selling (Qualitative Results)

USAID
FROM THE AMERICAN PEOPLE

Comparative Cost-Effectiveness of Four Supplementary Foods in Treating Moderate Acute Malnutrition in Children 6-59 Months In Sierra Leone

Section 2: Body Composition Sub-Study

A Report from the Food Aid Quality Review

Prepared by:

Devika Suri
Akriti Singh
Isabel Potani
Breanne Langlois
Stacy Griswold
Ye Shen
Ilana Cliffer
Kenneth Chui
Shelley Walton
Lindsey Green Ellis
Irwin Rosenberg
Patrick Webb
Beatrice Lorge Rogers

JANUARY 2021

This report was made possible by the generous support of the American people through the support of the United States Agency for International Development's Bureau for Humanitarian Assistance (USAID/BHA) and the legacy Office of Food for Peace (FFP) under the terms of Contract AID-OAA-C-16-00020, managed by Tufts University.

The authors have no conflict of interest to declare.

Recommended Citation

Suri, Devika; Singh, Akriti; Potani, Isabel; Langlois, Breanne; Griswold, Stacy; Shen, Ye; Cliffer, Ilana; Chui, Kenneth; Walton, Shelley; Green, Lindsey Ellis; Rosenberg, Irwin; Webb, Patrick; Rogers, Beatrice. 2020. *Comparative Cost-effectiveness of Four Supplementary Foods in Treating Moderate Acute Malnutrition in Children 6-59 Months in Sierra Leone- Secion 2: Body Composition Sub-study*, Report to USAID from the Food Aid Quality Review. Boston, MA: Tufts University.

For correspondence, contact:

Beatrice Lorge Rogers, PhD
Friedman School of Nutrition Science and Policy
Tufts University
150 Harrison Avenue
Boston, MA 02111
beatrice.rogers@tufts.edu
http://www.foodaidquality.org

ABBREVIATIONS AND ACRONYMS

BIA	Bioelectrical Impedance
C	Centigrade
Cm	Centimeter
CSB+	Corn-Soy Blend Plus
CWSB	Corn-Soy-Whey Blend
DD	Deuterium Dilution
D_2O	Deuterium Oxide
FBFs	Fortified Blended Foods
FFM	Fat-Free Mass
FM	Fat Mass
G	Gram
HAZ	Height-for-Age-Z-Score
IDRL	University Makeni Infectious Disease Research Laboratory
Kg	Kilogram
MAM	Moderate Acute Malnutrition
Ml	Mililiter
MUAC	Mid-Upper Arm Circumference
PHUs	Peripheral Health Units
RUSF	Ready-to-Use Foods
SAM	Severe Acute Malnutrition
SC+A	Super Cereal Plus Amylase
SNFs	Specialized Nutritious Foods
TBW	Total Body Water
W	Weight
WHZ	Weight-for-Height Z-score

TABLE OF CONTENTS
Section 2 : Body Composition Sub-study

ABSTRACT

Introduction: Anthropometric indicators such as mid-upper arm circumference (MUAC) and weight-for-height z-score (WHZ) are currently used to determine recovery from moderate acute malnutrition (MAM). Body composition measures of fat mass (FM) and fat-free mass (FFM), however, give more detail to the quality of weight gained and may be important for understanding longer term maintenance of recovery and risk of chronic disease later in life.

Objectives: To determine differential changes in children's body composition—fat-free mass (FFM) and fat mass (FM)—after 4 weeks of treatment for moderate acute malnutrition (MAM) with one of 4 four specialized nutritious foods (SNFs).

Methods: Body composition was assessed using the deuterium dilution technique at program enrollment and after 4 weeks of treatment among a subset of children enrolled in the cost-effectiveness study. Children were enrolled in 8 of the 29 peripheral health units in which the cost-effectiveness study was conducted. Total sample size was fixed at 200/arm for logistical and budgetary reasons. Changes in weight, FM, FFM and %FFM overall and by study arm were calculated; statistical significance was determined using t-tests and ANOVA (unadjusted).

Results: Among 312 subjects at enrollment, mean ± SD age was 12.9 ± 7.2 mos, weight was 6.5 ± 0.96 kg, FM was 1.3 ± 0.5 kg, FFM was 5.2 ± 0.95 kg, and %FFM was 80.1 ± 7.03. After 4 weeks of treatment, mean ± SD change in weight was 0.45 ± 0.39 kg ($P < 0.001$), change in FM was 0.07 ± 0.60 kg ($P = 0.005$), change in FFM was 0.37 ± 0.57 ($P < 0.001$), and change in %FFM was 0.26 ± 8.44(NS). Overall, weight gain consisted on average of 20.9% FM and 79.8% FFM, thus the %FFM was unchanged from enrollment to four weeks. Changes in FM and FFM were not significantly different across study arms. By study arm, mean ± SD changes in FM and FFM respectively, were: 0.05 ± 0.59 kg and 0.38 ± 0.54 kg in CSB+ w/oil; 0.15 ± 0.66 kg and 0.33 ± 0.64 kg in SC+A; 0.07 ± 0.63 kg and 0.37 ± 0.57 kg in CSWB w/oil; -0.01 ± 0.49 kg and 0.42 ± 0.49 kg in RUSF.

Conclusions: Over 4 weeks of treatment for MAM, children gained roughly 80% lean mass and 20% fat mass. This body composition is consistent with association between sustainability of recovery from MAM and healthier long-term recovery. Differential effects on body composition by type of SNF were not detected after 4 weeks of treatment.

1. Background and Objective

A number of different specialized nutritious foods (SNFs) have been used to treat children with moderate acute malnutrition (MAM), but it is recognized that supplementary feeding programs (SFPs) need to assess outcomes other than anthropometric indicators such as mid-upper arm circumference (MUAC) and weight-for-height Z-score (WHZ) as indicators of recovery (1,2). One such outcome is lean body mass or fat-free mass (FFM). While fat mass (FM) represents energy stores, fat-free mass (FFM) represents functional organs and tissues. Their proportion indicates how the body is prioritizing these needs (2).

Body composition, and not just body weight, has been shown to be important for survival in both healthy and sick children (3), and inadequate accumulation of either FM or FFM has been shown to be detrimental to health (4). Body composition in infancy can have lasting influence on adult noncommunicable disease risk (5). In particular, rapid and/or catch-up weight gain in early childhood has been associated with adiposity, obesity and non-communicable diseases later in life (5,6). Overall, however, there is limited understanding of the changes in body composition that occur during MAM and MAM treatment (7). At present, body composition reference data in young children is limited, and targets for absolute or proportional gain of FFM and FM in children while recovering from MAM are unclear (6,8).

In recognition of the possible implications of body composition for health at different stages of the life cycle, studies investigating body composition in MAM could help build evidence on the management of MAM and its impact on the global burden of disease. Studies on body composition and its relationship to anthropometric measures could also improve the understanding of how MUAC and WHZ capture changes in body composition during MAM treatment. The objectives of this sub-study were to investigate the effects of four SNFs on changes in FM and FFM among children during four weeks of treatment for MAM, and to compare these changes with anthropometric measures in order to expand understanding of factors influencing recovery and sustainability of recovery from MAM.

2. Methods

2.1 Study Design

This Body Composition Sub-Study was nested within the *Four Foods MAM Treatment Study,* a prospective, cluster-randomized, controlled clinical and cost-effectiveness trial assessing four SNFs to treat children 6-59 months with MAM, defined as MUAC ≥11.5 cm and <12.5 cm without bipedal edema. The sub-study was conducted in eight of the 29 peripheral health units (PHUs) in the *Four Foods MAM Treatment Study,* two per arm. Selection was purposively based on logistical needs and on the number of MAM cases to ensure that the target sample size was met within the allocated data collection time. The research team also considered the availability of sheltered and adequate space for study procedures to minimize the risk of sample contamination. Lastly, proximity of the data collection sites to the study office was a necessary logistical consideration. This was important because the study office was the only location in Pujehun, Sierra Leone that was accessible to the team and had consistent electrical power for a freezer to store the urine samples.

The PHUs selected from each study arm were:

- Corn-Soy Blend Plus (CSB+) w/oil: Blama Massaquoi and Makorma
- Super Cereal Plus with Amylase (SC+A): Sahn Malen and Taninahun
- Corn-Soy Whey Blend (CSWB) w/oil: Bandajuma and Kowama
- Ready-to-Use Supplementary Food (RUSF): Konia and Funyehun

2.2 Ethical Approvals

The sub-study was approved by the Tufts University Health Sciences Institutional Review Board and Sierra Leone Ethics and Scientific Review Committee. Written informed consent was obtained for all participants.

2.3 Participants and Sample Size

All subjects (beneficiary children and caregivers) at the selected PHUs were eligible to participate in the sub-study. A sample size of 200 per arm was determined, based on logistical/budgetary feasibility and the expected number of children to be recruited from the PHUs in the Sub-Study. The original primary outcome was change in the percentage of FFM after four weeks of treatment. (The main outcome on which results were compared was changes to absolute FM and FFM gain. See the "Limitations" section for further detail.) At the time of developing the study protocol, there were few studies of body composition in MAM using any of the food formulations under investigation that also used this outcome, so the researchers determined an estimated detectable effect using this maximum sample size.

To adjust for the cluster-randomized design, a conservative design effect of 1.115 was applied. This was because the outcome variable was continuous, with an expected low intra-class correlation, and therefore little variance between the clusters (i.e. PHUs) was expected. Furthermore, stratified analyses were planned based on age groups (6-23.9 months, and 24-36 months), thereby doubling the required sample size. As arms were capped at 200, power calculations were based on sample sizes of 89 per group: (i.e. 100/1.115). Using simple random sampling and a two-sided paired t-test, a sample size of 89 per group achieves 80% power to detect a mean of paired differences of 0.2 percentage points with an estimated standard deviation of differences of 0.5 and with a 0.05 significance level. This is consistent with a previous study (9) that examined similar foods, where change in percentage FFM after 12 weeks of treatment within each group ranged from -1.5 to 0.4 percentage points for fortified blended foods (FBFs), and -0.2 for RUSF.

2.4 Deuterium Dilution Method, Sample Collection and Analysis

To assess body composition, the research team used the deuterium dilution (DD) technique, which has been validated for assessing body composition in children with MAM (10) and is based on a two-component model of body composition (11). This technique requires the subject to ingest an oral dose of deuterium oxide (D_2O); their urine is collected approximately four hours later. Participants in the Body Composition Sub-Study were asked to return to the

clinic the *day after* collecting a ration from the PHU in order to have sufficient time for the body composition protocol (i.e. consuming the DD, waiting, and urine collection). This occurred at enrollment (baseline) and Week 4 (endline). This four-week time period was chosen so that all enrolled children would have the same exposure to the intervention, since children who graduated would not continue to receive the treatment food (those who graduated at two weeks would receive one more food ration which would bring them to a total of four weeks on the study food).

The body composition data collection team consisted of a research assistant (technical lead), two research assistant aides (in charge of recruiting participants, preparing and administering D_2O doses, and collecting urine samples), a driver (responsible for taking caregivers to and from the study site and their villages, as well as observing the caregiver's adherence to instructions), and a community health worker when one was available (assisted in locating caregiver villages, communicating instructions to caregivers and serving caregivers food).

Diluted D_2O was prepared using a 1:10 dilution of a 99% D_2O (Sigma Aldrich, US) with locally-purchased bottled water. Each dilution was approximately 1000 ml and was made of approximately 905 g of bottled drinking water and 90 g of D_2O. The dose was stored in a 1000-ml Nalgene bottle that was sealed with parafilm and stored in a freezer at the study office. Each stock solution lasted for approximately four weeks. Portions of the diluted dose and water used in the dilution were shipped for lab analysis. Since samples containing deuterium are susceptible to contamination—which can result in over or underestimation of body composition values—the first two weeks of the study were considered a pilot test of the DD technique under the field conditions in Pujehun. Additional samples of the dose, water, and urine for 16 study subjects from seven sites were sent for lab analysis and plausibility of the findings was confirmed.

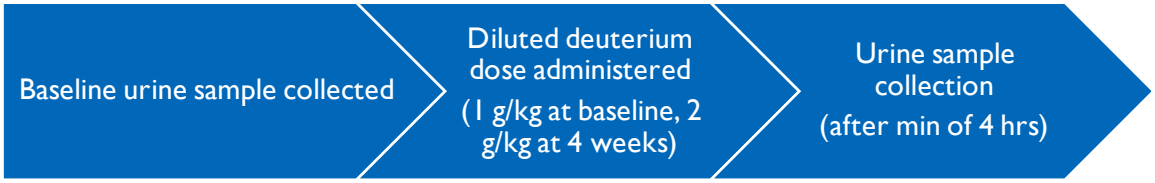
At baseline, participants were given a dose of diluted D_2O at 1 g/kg body weight. At endline (four weeks), the dose given was doubled to: 1) minimize natural variation in isotope enrichment in the post-dose urine sample; and 2) offset the effect of residual deuterium from the first dose on the deuterium enrichment at the four-week assessment. The doses were given orally using locally-purchased syringes, and subjects were not required to fast at any time before dose administration. The prescribed doses were weighed in grams to two decimal places at the study office and placed in syringes in sealed Ziploc bags.

Prepared doses were then transported in a cold box with gel ice packs. At the time of administration, the dose was weighed before and after administration, the difference between the two weights was the dose received by the participant. In cases when there were spills, the spills were recovered with pre,-weighed cotton balls. Losses were estimated from the difference in weights of the cotton balls before and after collecting the spill. When spills could not be successfully collected, this was noted in the participant's record. Other events that required noting on the day of D_2O dosing were vomiting, diarrhea, and any other occurrence that the study personnel thought might affect the outcome of the study. Subjects with greater than 10% spit-up or spilled deuterium solution were excluded from analysis.

After D_2O administration was complete, the participants were not allowed to take any food or drink for two hours, after which they were permitted to eat and drink water and breast milk under the watch of study personnel. While free intake of water was permitted, the amount of breast milk taken was estimated by weighing children before and after breast feeding. Older children were also served lunch, usually rice, local vegetables and meat. Younger children were served a rice porridge. Body water calculated from the DD method was adjusted for water coming from breast milk consumed during the four-hour protocol assuming breast milk has a water content of 87%. Metabolic water generated from non-aqueous food consumption during the four-hour period was considered to be negligible and was not adjusted.

Urine samples were collected before dosing with D_2O and after a minimum of four hours following D_2O administration (12), allowing the isotope to reach isotopic equilibrium (i.e. plateau value) in the body. The concentration of the excess deuterium, which is the difference between the pre-dose and post-dose samples, is then calculated. Urine was collected using PDC Healthcare-Assure Urine Collectors urine bags (ThermoFisher Scientific, US). Aliquots of approximately 1.5 ml of baseline and post-dose urine samples were transferred to airtight cryovials from the urine collection bags and then placed in cryoboxes. This was done at a urine collection station. The cryoboxes were stored and transported in an ice box to the study office for storage in a freezer at -20°C for at least four weeks, after which they were transferred to the University Makeni Infectious Disease Research Laboratory (IDRL) which was eight hours away. They were kept in -20°C freezer until shipment to the United States Department of Agriculture/Agricultural Research Service Children's Nutrition Research Center at Baylor College of Medicine in Houston, Texas for isotope ratio analyses. Refrigeration and freezing is not required but recommended for samples using the DD technique, due to potential mold and bacteria growth which would result in dilution of the isotopes in the urine samples (13).

Body Composition Data Collection Procedure



The deuterium content of the urine samples, the diluted dose and water used to prepare it were analyzed in duplicate at the Baylor lab using continuous-flow isotope ratio mass spectrometry (14). The lab reported the deuterium content of the diluted D_2O dose, the water used to prepare the diluted dose and the pre- and post-dose urine samples. The deuterium dilution space, as well as FFM, FM, and percent FM with adjustment, if necessary, of missing dose and/or milk intake during the four-hour equilibration period were calculated as shown in **Appendix I.**

2.5 Data Management

The body composition study used a data collection form at enrollment and at four weeks, which was then uploaded to a KoboCollect[1] data collection tool kit using a laptop computer at each study site. All demographic and anthropometric data were collected for the *Four Foods MAM Treatment Study*. However, for some children who graduated at two weeks (or did not show up at the fourth week visit), MUAC, weight and length were collected as per study protocol at the four-week time point by the Body Composition sub-study team. Prior to sending samples to the lab, relevant participant information was exported from the KoBoCollect online database to minimize typographical errors. A clean data file in Excel format along with the samples was sent to the analyzing lab for deuterium measurements.

Unphysiological total body water (TBW) values due to errors in dose weight and dose spillage (>10%) were excluded from the dataset (n = 92). Since the quality control of the mass spectrometric measurements is well established (13), errors at various steps of the study most likely are the primary causes of implausible TBW values. Data management and cleaning were conducted with Stata 15 and Microsoft Excel 2017.

2.6 Research Questions and Hypotheses

After four weeks of treatment:

1. What are the changes in body composition (specifically FM and FFM) of 6-59 month-old children with MAM after four weeks of treatment? (Exploratory, no *a priori* hypothesis)
2. What is the relation of FM and FFM to MUAC, WHZ and recovery?
   Hypothesis: Changes in body composition will be correlated with changes in MUAC, WHZ and recovery.
3. Are there differences in body composition changes among the four SNFs?
   Hypothesis: The four food supplements will have different effects on the body composition of the 6-59 month-old children with MAM.

Body Composition Measures:

- TBW, kg
- FFM, kg
- FM, kg
- FFM percent (FFM/body weight x 100)
- FFM index (FFM kg/height^2)

Anthropometric Measures and Covariates (from study clinic data):

- Weight, kg
- MUAC, cm (*Four Foods MAM Treatment Study* main outcome, used to determine recovery from MAM)
- Weight-for-height z-score
- Age of child at enrollment, in months

[1] Kobotoolbox: https://www.kobotoolbox.org/

- Sex of child
- Wealth quintile
- Whether the child graduated into the MAM program from severe acute malnutrition (SAM) treatment (0=Not SAM start; 1=SAM start)
- Whether the child recovered from MAM within 12 weeks (0=Did not recover; 1= Recovered in 12 weeks)

2.7 Statistical Analysis

The dataset used for the Body Composition Sub-Study analysis consisted of body composition measures (listed below) plus anthropometric and demographic variables from the *Four Foods MAM Treatment Study* clinic cards.

To answer the first research question, descriptive statistics for baseline and outcome values of body composition and anthropometry for the included subjects were tabulated. The average changes in these variables were calculated, as well as the average percent of weight gain that was from FM and FFM.

To address the second research question and hypothesis, the research team calculated Pearson correlations among the changes in FM, FFM, MUAC and WHZ. The research team further estimated mean changes in FFM, FM and weight by whether a child recovered by 12 weeks and whether that recovery was sustained at the 4-week follow-up visit, using margins from logistic regressions of recovery or sustained recovery (binary) on each outcome, adjusted for baseline FFM, age, sex, baseline WHZ and wealth quintile.

To address the third research question and hypothesis, the research team ran linear regression models by study arm on change in FFM and FM, controlling for baseline body composition (FFM or FM), age, sex, baseline WHZ and wealth quintile. Statistical differences among the study arms was assessed using a Wald test. Mixed models were used to adjust for clustering at the PHU levels, but these results showed a negligible design effect, and thus, the final models did not adjust for clustering. Models were assessed for goodness-of-fit, collinearity, and potentially-influential outliers. Adjusted outcome values for change over four weeks were obtained using predicted margins. Data analysis procedures were conducted with Stata 15 and Microsoft Excel 2017.

3. Results

3.1 Study Population and Characteristics

A total of 578 children were recruited and their caregivers consented to participation in the body composition study. Of these, 516 children completed the baseline body composition procedure and 407 completed the four-week follow-up procedure. Reasons for not completing the four-week measure included loss to follow up (38), progression to SAM (50), unsuccessful or missing samples (11) and death (4). Six subjects were dropped from the *Four Foods MAM Treatment Study* dataset due to data collection issues. Once the urine samples were

analyzed and the body composition measures had been calculated, 92 subjects were removed from the analysis dataset due to methodology issues (see "2.5 Data Management" and "2.7 Statistical Analysis" sections above). Two values were identified as influential outliers and removed from analysis. A total of 312 subjects were included in the final analysis dataset.

Baseline descriptive statistics for the 312 subjects included in the Body Composition Sub-Study analysis can be found in **Table 1**. Although the age range for the program spanned 6-59 months, children were on average about 12 months of age, and there were slightly more females than males. A quarter of subjects had graduated into the MAM treatment program from SAM treatment. The wealth quintiles were significantly different among the four study arms.

Table 1: **Baseline Descriptive Statistics of Subjects Included in Body Composition Analysis, *n*=312**

	Total	CSB+ w/ oil	SC+A	CSWB w/ oil	RUSF	*P*-value[3]
n	312	84	94	68	66	
Age, months	12.19±7.18[1]	11.98±6.64	12.69±7.80	12.45±7.20	11.46±6.97	0.733
Females	173 (55%)[2]	49 (58%)	55 (59%)	35 (51%)	34 (52%)	0.685
Previous SAM	78 (25%)	23 (28%)	23 (24%)	15 (22%)	17 (26%)	0.880
HAZ	-2.72±1.15	-2.73±1.13	-2.73±1.04	-2.84±1.26	-2.58±1.19	0.634
WHZ	-1.70±0.74	-1.63±0.69	-1.76±0.74	-1.85±0.76	-1.57±0.75	0.105
Wealth quintile						<0.001
Lowest	69 (22%)	22 (26%)	11 (12%)	17 (25%)	19 (29%)	
Low	52 (17%)	9 (11%)	13 (14%)	9 (13%)	21 (32%)	
Middle	62 (20%)	16 (19%)	23 (25%)	10 (15%)	13 (20%)	
High	71 (23%)	20 (24%)	21 (23%)	18 (26%)	12 (18%)	
Highest	56 (18%)	17 (20%)	24 (26%)	14 (21%)	1 (2%)	
Recovered within 12 weeks	229 (73%)	61 (73%)	70 (74%)	51 (75%)	47 (71%)	0.954

[1] Mean ± SD, all such values
[2] Frequency (percent), all such values
[3] ANOVA tests were used for continuous variables; chi-square tests were used for categorical variables
Abbreviations: CSB+, Corn-Soy Blend Plus; CSWB, Corn-Soy-Whey Blend; HAZ, height-for-age z-score; RUSF, ready-to-use supplementary food; SAM, severe acute malnutrition; SC+A, Super Cereal plus Amylase; WHZ, weight-for-height z-score

Research Question 1: What are the changes in body composition in children with MAM after four weeks of treatment?

Baseline anthropometric and body composition measures are shown in **Table 2**. There were no significant differences in baseline measures among the study arms. Overall, changes in anthropometric and body composition measures over four weeks of treatment were highly variable among study subjects. Weight gain ranged from -0.65 kg to 2.1 kg, change in FFM ranged from -1.5 kg to 2.2 kg, and change in FM ranged from -1.8 kg to 1.6 kg (**Figure 1**).

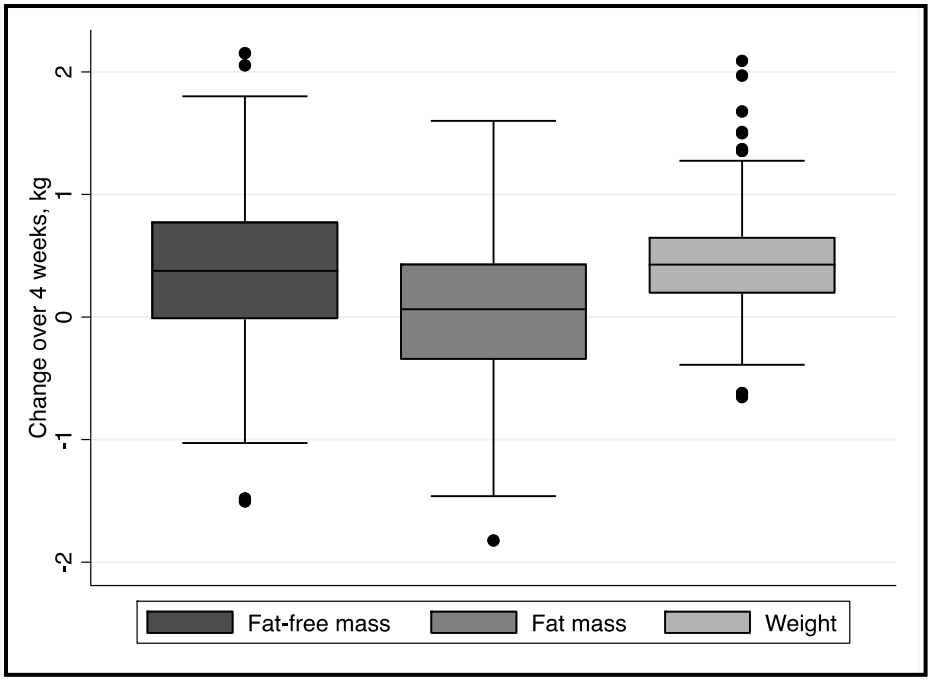
Percent FFM was even more varied, with changes over four weeks ranging from -21.1 to 22.6 percentage points. Unadjusted mean changes in anthropometric and body composition measures over four weeks of treatment are also shown in **Table 2**. On average, children gained weight over the four weeks of treatment; approximately 82% of the weight gained during this period was FFM (a slightly higher proportion of FFM compared to average baseline FFM of 80.1%).

Table 2: Unadjusted Anthropometric and Body Composition Measures at Baseline and Changes After Four Weeks of Treatment for MAM, *n* = 312

	Total	CSB+ w/ oil	SC+A	CSWB w/ oil	RUSF	*P*-value[2]
n	312	84	94	68	66	
Baseline						
Weight, kg	6.53±0.961	6.55±1.09	6.57±0.98	6.52±0.78	6.47±0.95	0.915
WHZ	-1.70±0.74	-1.63±0.69	-1.76±0.74	-1.85±0.76	-1.57±0.75	0.105
MUAC, cm	11.97±0.27	12.01±0.27	11.98±0.26	11.94±0.27	11.93±0.28	0.264
Total body water, kg	4.16±0.71	4.15±0.64	4.19±0.81	4.16±0.65	4.11±0.72	0.914
FFM, kg	5.24±0.95	5.23±0.86	5.29±1.07	5.25±0.87	5.18±0.96	0.901
FM, kg	1.29±0.48	1.33±0.52	1.28±0.49	1.27±0.49	1.29±0.43	0.871
FFM %	80.11±7.03	80.00±6.39	80.17±7.48	80.42±7.57	79.84±6.71	0.968
FFM index	11.62±1.15	11.64±1.18	11.57±1.20	11.62±1.16	11.66±1.04	0.968
Change after 4 weeks[4]						
Weight, kg	0.45±0.39	0.43±0.39	0.48±0.41	0.44±0.38	0.41±0.38	0.696
WHZ[3]	0.30±0.61	0.24±0.57	0.35±0.57	0.42±0.65	0.16±0.64	0.154
MUAC, cm [4]	0.32±0.46	0.29±0.49	0.32±0.41	0.37±0.49	0.30±0.48	0.827
Total body water, kg	0.29±0.45	0.29±0.43	0.26±0.51	0.28±0.45	0.33±0.39	0.791
FFM, kg	0.37±0.57	0.38±0.54	0.33±0.64	0.37±0.57	0.42±0.49	0.807
FM, kg	0.07±0.60	0.05±0.59	0.15±0.66	0.07±0.63	-0.01±0.49	0.415
FFM %	0.26±8.44	0.34±8.14	-0.65±9.03	0.28±8.81	1.46±7.54	0.492
FFM index[5]	0.49±1.28	0.57±1.21	0.43±1.49	0.42±1.19	0.54±1.18	0.899

[1] Mean ± SD, all such values
[2] ANOVA tests used to determine differences among the 4 study arms
[3] *n*=257 due to missing values at 4 weeks
[4] *n*=250 due to missing values at 4 weeks
[5] *n*=234 due to missing values at 4 weeks
Abbreviations: CSB+, Corn-Soy Blend Plus; SC+A, Super Cereal plus Amylase; CSWB, Corn-Soy-Whey Blend; RUSF, ready-to-use supplementary food; WHZ, weight-for-height z-score; MUAC, mid-upper arm circumference; FFM, fat-free mass; FM, fat mass

Figure 1: Changes in FFM, FM and Weight Over Four Weeks, *n* = 312



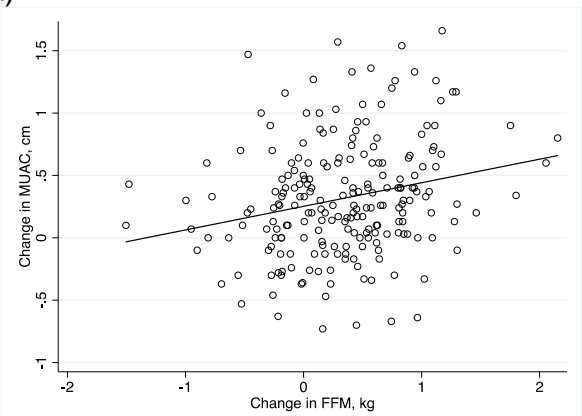
Research Question 2: What is the relation of FM and FFM to MUAC, WHZ and recovery?

The results of this study supported the hypothesis that changes in body composition and changes in anthropometric indicators were correlated. While the correlations were statistically significant, however, they were fairly weak. Scatterplots of changes in MUAC versus changes in FFM or FM are shown in **Figure 2a** and changes in WHZ versus changes in FFM or FM are shown in **Figure 2b**. Pearson correlations and adjusted R-squared values were similar.

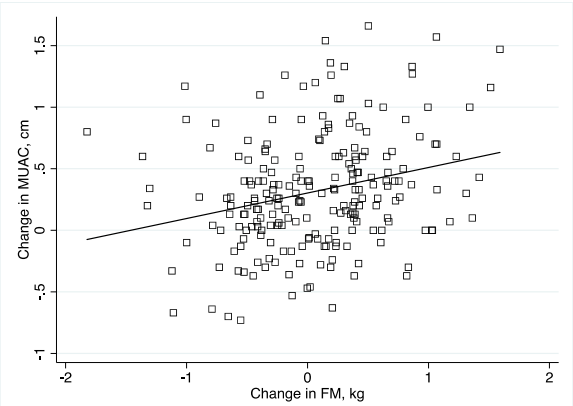
Gains in FFM, FM and weight in the first four weeks of treatment for MAM were significantly higher among children who recovered within the full 12 weeks of treatment (**Figure 3**). Children who recovered within 12 weeks gained an average of 0.43 kg FFM during the first 4 weeks of treatment, while those who did not recover by 12 weeks gained an average of 0.22 kg FFM ($P = 0.002$). Similarly, children who recovered within 12 weeks gained an average of 0.14 kg FM during the first four weeks of treatment, while those who did not recover by 12 weeks *lost* an average of 0.13 kg FM ($P < 0.001$). In terms of proportions of FFM and FM, children who recovered gained in a similar proportion to their starting body composition, on average, while children who did not recover gained proportionately more FFM than FM, on average, resulting in a higher percent FFM at endline. The pattern was similar for children who recovered and whether or not they sustained that recovery at their one-month follow-up or relapsed to MAM (**Figure 4**). Children who sustained recovery tended to have gained more FFM, FM and weight over the first four weeks of treatment, though only weight gain was significantly different between the groups ($P<0.001$).

Figure 2: Changes in FFM and FM Over Four Weeks Versus Change in MUAC (a) and WHZ (b), *n*=218 and *n*=225, Respectively.

a)

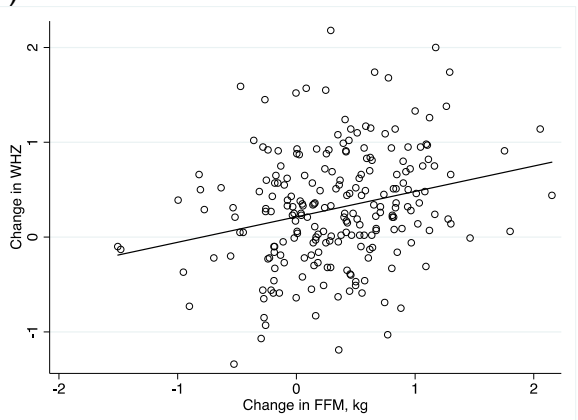


Adjusted R-squared: 0.05
Pearson correlation: 0.23 (P<0.001)

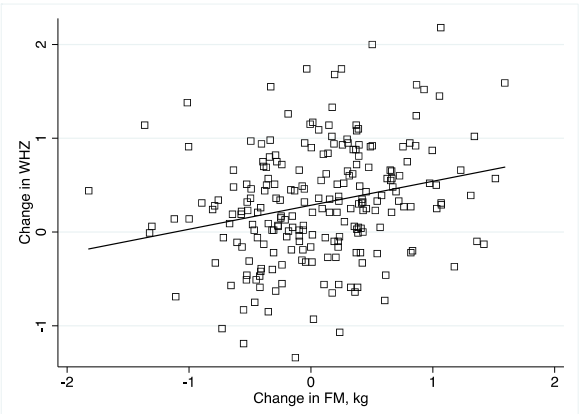


Adjusted R-squared: 0.06
Pearson correlation: 0.26 (P<0.001)

b)



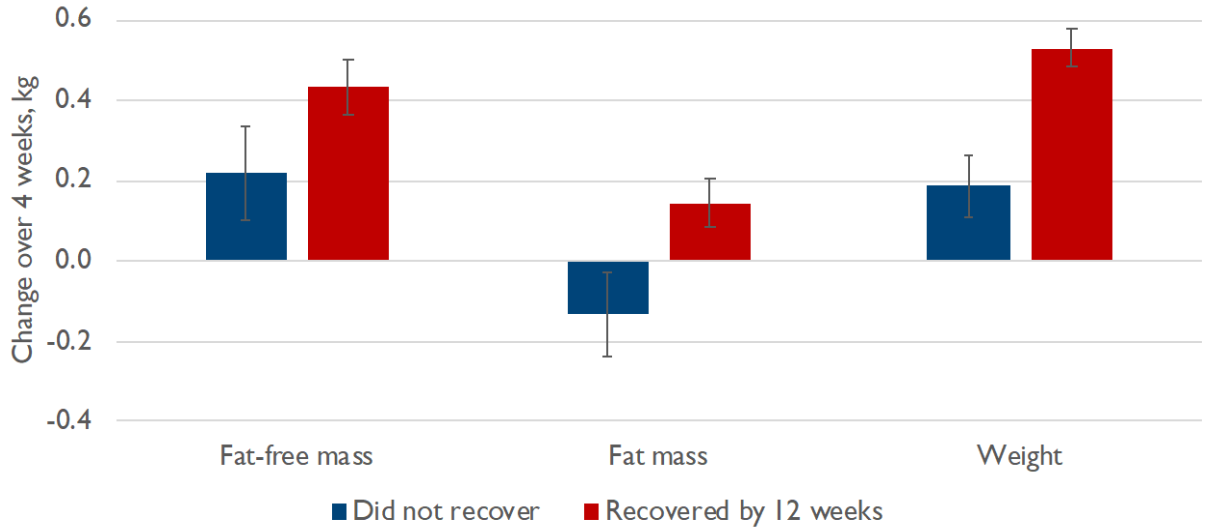
Adjusted R-squared: 0.06
Pearson correlation: 0.25 (P<0.001)



Adjusted R-squared: 0.06
Pearson correlation: 0.24 (P<0.001)

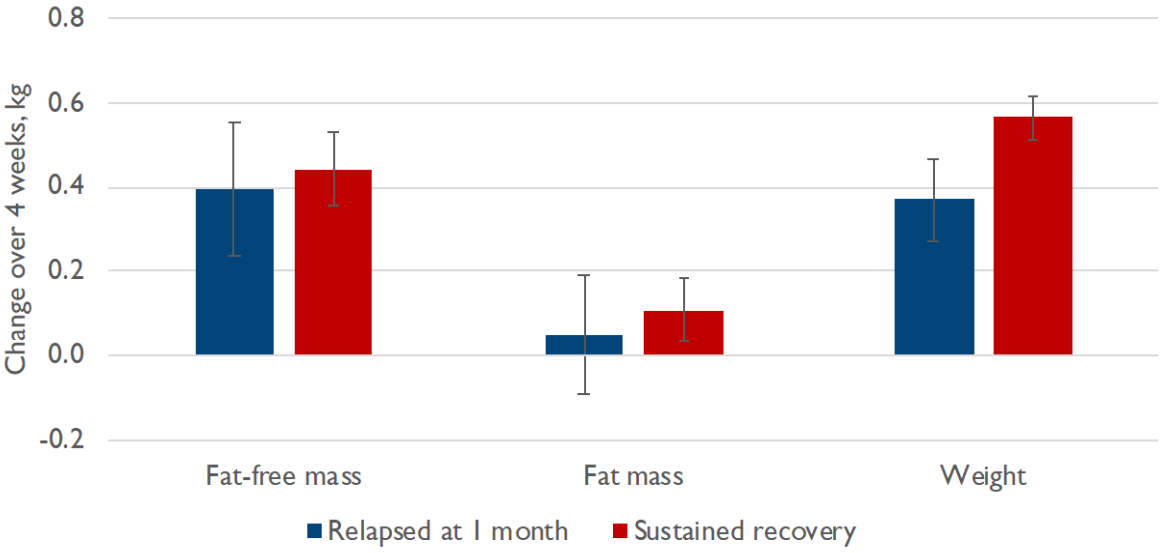
Abbreviations: FFM, fat-free mass; FM, fat mass; MUAC, mid-upper arm circumference; WHZ, weight-for-height z-score

Figure 3: Changes in FFM, FM and Weight Within First Four Weeks of Treatment, Comparing Children Who Recovered Within 12 Weeks Of Treatment for MAM and Those Who Did Not[1], *n*=312



[1] Difference in change in FFM, FM and weight by recovery status: $P = 0.002$, $P < 0.001$ and $P < 0.001$, respectively; recovered $n = 229$, did not recover $n = 83$
Error bars represent 95% confidence intervals
Values are predicted margins estimates using linear models adjusted for baseline value of measure, study arm, sex, baseline WHZ and wealth quintile
Abbreviations: FFM, fat-free mass; FM, fat mass; MAM, moderate acute malnutrition

Figure 4: Changes in FFM, FM and Weight Within First Four Weeks of Treatment, Comparing Children Who Sustained Recovery for One Month After Program Exit With Those Who Relapsed to MAM, *n* = 217[1]

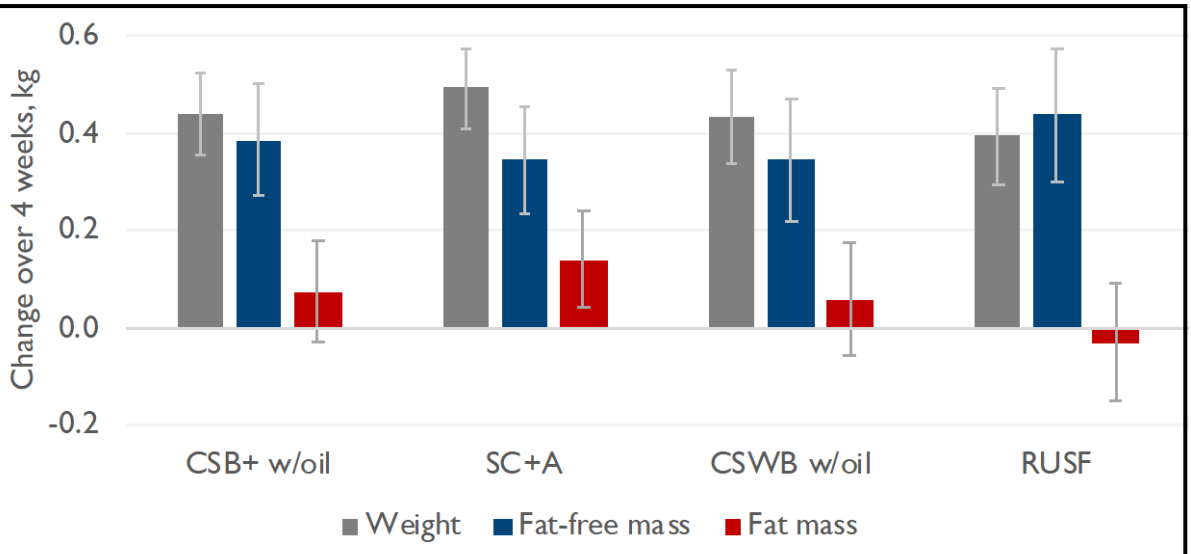


[1] Difference in change in FFM ($P = 0.605$), FM ($P = 0.479$) and weight ($P < 0.001$) by recovery status: relapsed at 1 month, $n = 50$; sustained recovery $n = 169$; and missing $n = 93$
Error bars represent 95% confidence intervals
Values are predicted margins estimates using linear models adjusted for baseline value of measure, study arm, sex, baseline WHZ and wealth quintile
Abbreviations: FFM, fat-free mass; FM, fat mass.

Research Question 3: Are there differences in body composition changes among the four SNFs?

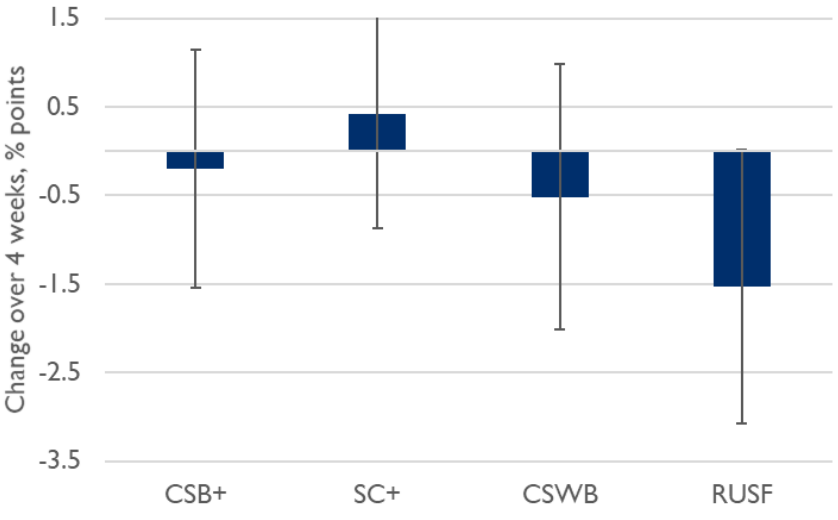
The results of this study did not find differential changes in body composition by study food. There were no differences in change in anthropometry or body composition across the groups (**Table 2**). An adjusted analysis of changes in weight, FM, FFM and %FM also showed no significant difference (**Figure 5** and **Figure 6**).

Figure 5: Changes in Weight, Fat-Free Mass and Fat Mass by Study Arm After Four Weeks of Treatment for MAM by Study Arm, *n* = 312.[1]



[1] Models adjusted for baseline outcome measure, age, sex, baseline WHZ and wealth quintile
Abbreviations: MAM, moderate acute malnutrition; CSB+, Corn-Soy Blend Plus; SC+A, Super Cereal plus Amylase; CSWB, Corn-Soy Whey Blend; RUSF, ready-to-use supplementary food.

Figure 6: Changes in Percent of Fat Mass by Study Arm After Four Weeks of Treatment for MAM by Study Arm, *n* = 312[1]



Models adjusted for baseline outcome measure, age, sex, baseline WHZ and wealth quintile
Abbreviations: MAM, moderate acute malnutrition; CSB+, Corn-Soy Blend Plus; SC+A, Super Cereal plus Amylase; CSWB, Corn-Soy-Whey Blend; RUSF, ready-to-use supplementary food.

4. Limitations and Challenges

4.1 Limitations

Due to the methodological challenges described above, the final analyzable sample size was reduced by almost a quarter. Because the proportion of dropped subjects was similar among the study arms, it is not likely that this biased the results. However, it did reduce the statistical power to detect any potential differences among the study arms in terms of body composition changes.

4.2 Challenges

The research team faced a number of challenges during the Body Composition Sub-Study. First, the oral diluted deuterium dose was poorly accepted by the partcipants possibly due to reduced appetite (common in malnourished children), and participants often refused, spit up or spilled the deuterium solution dose. This loss was estimated using cotton balls, which is a poor estimator of deuterium loss but the only practical method available at present. Second, urine sampling resulted in a high time burden on the participants and caregivers due to the long equilibration time. Frustration was observed as caregivers would attempt to force the participants to drink more water to facilitate the post-dose urine sample collection. Children with MAM are often dehydrated which reduces the urge to urinate.

5. Summary of Findings

This study found that body composition (FM and FFM) changes measured by deuterium dilution provided additional information to the outcomes of recovery and sustainability of recovery in children treated for MAM. In summary:

1. Over four weeks of treatment, children on average gained mostly FFM.
2. FM and FFM were weakly correlated with anthropometric measures MUAC and WHZ.
3. Children who recovered from MAM by 12 weeks were already showing a different pattern of weight gain at four weeks compared to children who did not recover by 12 weeks: on average, higher gains in FM and FFM, and no overall change to the proportion of FM and FFM (whereas children who did not recover had an increase in the proportion of FFM to FM, on average).
4. Similarly, children who sustained recovery for one month after graduation showed greater gains in both FM and FFM during the first four weeks of treatment than those who did not.
5. Changes in FM, FFM, %FFM, and weight gain were similar across the four SNFs.

The findings of this study are comparable to other studies that found children gained mainly FFM during treatment of MAM. A nine-month supplementary feeding study in Cambodia looked at changes in FM and FFM among 419 children starting at six months of age and found an average increase in weight of 1.73 kg, with a 1.96 kg increase in FFM and a 0.21 kg decrease in FM; unfortunately, the authors did not report % change in FFM (15). A study in Burkina Faso found that 93.5% (95% CI 89.6, 97.4) of the weight gain comprised of FFM (16), a bit higher than the finding of 82% FFM of the weight gain in this study. A study in Mali compared body composition changes in children 6-35 months with MAM treated for 12 weeks with one of four supplementary foods. Body weight increased on average 1.02 ± 0.54 kg, FFM 0.69 ± 0.53 kg (67.6% of weight gain), and FM 0.31 ± 0.54 kg (30.4% of weight gain). Stratifying by recovery, they saw a similar pattern to these results, with children who recovered gaining more in FFM and FM than those who did not recover (FFM 0.81 ± 0.50 kg, FM 0.44 ± 0.52 kg and FFM 0.50 ± 0.56 kg, FM 0.09 ± 0.51 kg, respectively) (17,18). These findings contribute to evidence that may alleviate concers that "unhealthy weight gain" may occur during treatment of wasting with lipid-based or other supplementary foods (19).

While there appeared to be slight differences in the trends for FFM and FM gain among study arms, these were not statistically significant. This could be due to the relatively short duration between measures (four weeks), combined with the variability in both measurement of body composition and children's responses to the intervention. This study was different from previous studies in that it was the shortest duration between baseline and endline body composition measures, i.e. four weeks compared with 12 weeks at the minimum in other studies. (The choice of four weeks was made for logistical reasons, to have comparable study food exposure; see the "2. Methods" section.)

Other studies also restricted their maximum age to two or three years or followed a cohort of children of the same age over time. Similar studies did find statistically significant differences between study foods: a supplementary food trial in Burkina Faso compared children 6-23 months of age with MAM (n = 1,609) treated for 12 weeks and found that children treated with a lipid-based nutrient supplement (similar to RUSF) had significantly more gains in FFM index (but not FFM g) compared with CSB (16). However, a study in Mali comparing body composition changes in children 6-35 months with MAM (n = 289) treated for 12 weeks with one of four supplementary foods, found a higher increase in FM in the RUSF group compared to CSB++, although there was no difference in FM gains as a percent of body weight among the groups (17,18). While this study found no differences among the study arms, future studies with higher sample size and longer duration (i.e. higher statistical power) are warranted.

References

1. Green LE, Cliffer IR, Suri DJ, Caiafa KR, Rogers BL, Webb PJR. Advancing Nutrition in the International Food Assistance Agenda: Progress and Future Directions Identified at the 2018 Food Assistance for Nutrition Evidence Summit. *Food Nutr Bull* [Internet]. 2019;037957211987171. Available from: http://journals.sagepub.com/doi/10.1177/0379572119871715
2. Friis H, Michaelsen KF, Wells JC. Choice of Design and Outcomes in Trials among Children with Moderate Acute Malnutrition. *Food Nutr Bull* [Internet]. 2015;36:S35–40. Available from: http://journals.sagepub.com/doi/10.1177/15648265150361S106
3. Radhakrishna KV, Kulkarni B, Balakrishna N, Rajkumar H, Omkar C, Shatrugna V. Composition of weight gain during nutrition rehabilitation of severely under nourished children in a hospital based study from India. *Asia Pac J Clin Nutr*. 2010;19:8–13.
4. Henry CJ. Body mass index and the limits of human survival. *Eur J Clin Nutr*. 1990;44:329–35.
5. Wells JCK. Using Body Composition Assessment to Evaluate the Double Burden of Malnutrition. *Ann Nutr Metab* [Internet]. 2019;75:103–8. Available from: https://www.karger.com/Article/FullText/503666
6. Kristjansson E, Francis DK, Liberato S, Benkhalti Jandu M, Welch V, Batal M, Greenhalgh T, Rader T, Noonan E, Shea B, et al. Food supplementation for improving the physical and psychosocial health of socio-economically disadvantaged children aged three months to five years. *Cochrane Database Syst Rev* [Internet]. John Wiley & Sons, Ltd; 2015 [cited 2018 Mar 26]; Available from: http://doi.wiley.com/10.1002/14651858.CD009924.pub2
7. Bhutta Z, Yackobovitch-Gavan M. Malnutrition and Catch-Up Growth during Childhood and Puberty. *Food and Nutrition Bulletin* [Internet]. 2015. p. 120–7. Available from: https://www.karger.com/Article/FullText/441823
8. Wells JCK. Toward Body Composition Reference Data for Infants, Children, and Adolescents. *Adv Nutr*. 2014;
9. McDonald C, Ackatia Armah R, Doumbia S, Brown K. The effect of dietary supplementation on the change in body composition of young Malian children with moderate acute malnutrition (620.2) . *FASEB J*. 2014;28.
10. Fabiansen C, Yaméogo CW, Devi S, Friis H, Kurpad A, Wells JC. Deuterium dilution technique for body composition assessment: resolving methodological issues in children with moderate acute malnutrition. *Isotopes Environ Health Stud* [Internet]. 2017;53:344–55. Available from: https://www.tandfonline.com/doi/full/10.1080/10256016.2017.1295043
11. Shepherd JA, Heymsfield SB, Norris SA, Redman LM, Ward LC, Slater C. Measuring body composition in low-resource settings across the life course. *Obesity* [Internet]. 2016;24:985–8. Available from: http://doi.wiley.com/10.1002/oby.21491
12. Schoeller DA, van Santen E, Peterson DW, Dietz W, Jaspan J, Klein PD. Total body water measurement in humans with 18O and 2H labeled water. *Am J Clin Nutr*. 1980;33:2686–93.
13. IAEA. Introduction to Body Composition Assessment Using the Deuterium Dilution Technique with Analysis of Urine Samples by Isotope Ratio Mass Spectrometry. IAEA

Hum Heal Ser no 13 [Internet]. *International Atomic Energy Agency*; 2011;84. Available from: http://www-pub.iaea.org/books/iaeabooks/8370/Introduction-to-Body-Composition-Assessment-Using-the-Deuterium-Dilution-Technique-with-Analysis-of-Urine-Samples-by-Isotope-Ratio-Mass-Spectrometry

14. Wong WW, Clarke LL. A Hydrogen Gas-Water Equilibration Method Produces Accurate and Precise Stable Hydrogen Isotope Ratio Measurements in Nutrition Studies. *J Nutr*. 2012;
15. Skau JK, Touch B, Chhoun C, Chea M, Unni US, Makurat J, Filteau S, Wieringa FT, Dijkhuizen MA, Ritz C, et al. Effects of animal source food and micronutrient fortification in complementary food products on body composition, iron status, and linear growth: A randomized trial in Cambodia. *Am J Clin Nutr* [Internet]. Oxford University Press; 2015;101:742–51. Available from: https://academic.oup.com/ajcn/article/101/4/742/4564489
16. Fabiansen C, Yameogo CW, Iuel-Brockdorf AS, Cichon B, Rytter MJH, Kurpad A, Wells JC, Ritz C, Ashorn P, Filteau S, et al. Effectiveness of food supplements in increasing fat-free tissue accretion in children with moderate acute malnutrition: A randomised 2 x 2 x 3 factorial trial in Burkina Faso. *PLoS Med.* 2017/09/12. 2017;14:e1002387.
17. McDonald C, Armah RA, Doumbia S, Brown K. The effect of dietary supplementation on the change in body composition of young Malian children with moderate acute malnutrition (620.2). *FASEB J.* 2014;
18. McDonald CM, Ackatia-Armah RS, Doumbia S, Kupka R, Duggan CP, Brown KH. Percent Fat Mass Increases with Recovery, But Does Not Vary According to Dietary Therapy in Young Malian Children Treated for Moderate Acute Malnutrition. *J Nutr.* Oxford University Press; 2019;149:1089–96.
19. Organization WH. Supplementary foods for the management of moderate acute malnutrition in children aged 6–59 months [Internet]. *e-Library of Evidence for Nutrition Actions (eLENA)*. [cited 2019 Sep 16]. Available from: https://www.who.int/elena/titles/food_children_mam/en/
20. Butte NF, Hopkinson JM, Wong WW, Smith EO, Ellis KJ. Body composition during the first 2 years of life: an updated reference. *Pediatr Res.* 2000;47:578–85.

Appendix 1: Estimation of Body Composition

In order to calculate the deuterium dilution space (N) the following equation was used:

$$N = (T \cdot A/a) \cdot [(Ed–Et)/(Es–Ep)]$$

In the equation, T is the mass of water used to dilute the D_2O dose in grams, A is the mass in grams of the ingested dose in the field, a is the mass in grams of the dose that is diluted in the lab for mass spectrometry analysis, Ed is the isotopic abundance (ppm) in the diluted dose solution, and Et is the isotope abundance in ppm of the water used to dilute the dose, Es is the isotope abundance of the post-dose urine sample in ppm and Ep is the isotope abundance in ppm of the pre-dose urine sample. The dilution space (N) is then divided by a proton exchange factor of 1.040 to account for non-aqueous exchange of ^2H in order to convert it to total body water (TBW). FFM is calculated from TBW using age and gender specific appropriate hydration factors (H) (20): FFM = TBW / H

Below are the best fit equations to calculate the hydration factors:

Girls: $H = (-0.0003 \times Age^3 + 0.027 \times Age^2 – 0.704 \times Age + 84.064)/100$
Boys: $H = (-0.001 \times Age^3 + 0.037 \times Age^2 – 0.710 \times Age + 83.764)/100$

FM is then calculated as the difference between body weight (W) and FFM: FM = W - FFM

USAID
FROM THE AMERICAN PEOPLE

Comparative Cost-Effectiveness of Four Supplementary Foods in Treating Moderate Acute Malnutrition in Children 6-59 Months in Sierra Leone

Section 3: Environmental Enteric Dysfunction Sub-Study

A Report from the Food Aid Quality Review

Prepared by:

Akriti Singh
Breanne Langlois
Stacy Griswold
Ye Shen
Ilana Cliffer
Isabel Potani
Devika Suri
Kenneth Chui
Shelley Walton
Lindsey Green Ellis
Irwin Rosenberg
Patrick Webb
Beatrice Lorge Rogers

JANUARY 2021

This report was made possible by the generous support of the American people through the support of the United States Agency for International Development's Bureau for Humanitarian Assistance (USAID/BHA) and the legacy Office of Food for Peace (FFP) under the terms of Contract AID-OAA-C-16-00020, managed by Tufts University.

The authors have no conflict of interest to declare.

Recommended Citation

Singh, Akriti; Langlois, Breanne; Griswold, Stacy; Shen, Ye; Cliffer, Ilana; Suri, Potani, Isabel; Devika; Chui, Kenneth; Walton, Shelley; Green, Lindsey Ellis; Rosenberg, Irwin; Webb, Patrick; Rogers, Beatrice. 2020. *Comparative Cost-effectiveness of Four Supplementary Foods in Treating Moderate Acute Malnutrition in Children 6-59 Months in Sierra Leone- Section 3: Environmental Enteric Dysfunction Sub-study*, Report to USAID from the Food Aid Quality Review. Boston, MA: Tufts University.

For correspondence, contact:

Beatrice Lorge Rogers, PhD
Friedman School of Nutrition Science and Policy
Tufts University
150 Harrison Avenue
Boston, MA 02111
beatrice.rogers@tufts.edu
http://www.foodaidquality.org

ABBREVIATIONS AND ACRONYMS

AAT	Alpha-I-Antitrypsin
EED	Environmental Enteric Dysfunction
EPS	EED Protein Score
GIS	Gut Inflammation Score
GAPDH	Glyceraldehyde 3-Phosphate Dehydrogenase
GDS	Gut Defense Score
GSS	Gut Structure Score
SNF	Supplementary Nutritious Foods
LMER	Lactulose:Mannitol Excretion Ratio
MAL-ED	Malnutrition and Enteric Disease
MAM	Moderate Acute Malnutrition
MPO	Myeloperoxidase
mRNA	Messenger Ribonucleic Acid
MUAC	Mid-Upper Arm Circumference
NEO	Neopterin
PCoA	Principle Coordinates Analysis
PERMANOVA	Permutational Multivariate Analysis of Variance
PHU	Peripheral Health Unit
PPB	Project Peanut Butter
SD	Standard Deviation
WASH	Water, Sanitation, and Hygiene

TABLE OF CONTENTS

Section 3: Environmental Enteric Dysfunction Sub-study

ABSTRACT

Introduction: Environmental enteric dysfunction (EED), an impairment of the structure and function of the small intestine, may be involved in processes that lead to undernutrition, including moderate acute malnutrition (MAM). However, little is known about the role of EED during MAM treatment with specialized nutritious foods (SNFs) or about factors associated with EED during MAM. Measuring EED in the field is also challenging, since current biomarkers measure specific domains of EED, and some of the newer measures have yet to be validated against standard measures or compared with each other.

Objectives: The objectives of the EED sub-study were to examine 1) whether EED modifies the effects of SNFs on recovery from MAM (reaching a mid-upper arm circumference greater than or equal to 12.5cm), 2) if there is an improvement in EED over four weeks of treatment with SNFs, 3) how EED biomarkers compare with each other and with measures of intestinal inflammation and damage, and 4) the association between water, sanitation, and hygiene (WASH) practices and EED, as well as 5) the association between the microbiota profile of MAM children and their varying levels of EED.

Methods: At eight of the peripheral health units of the *Four Foods MAM Treatment Study* in Pujehun district of Sierra Leone, EED was assessed at enrollment (n=484) using the lactulose: mannitol (L:M) test, fifteen fecal host messenger RNA (mRNA) transcripts[1] and three fecal host proteins[2]. Fecal host mRNA transcripts were also assessed after 4 weeks of supplementation. Factor analysis was used to develop three scores using the fecal host mRNA transcripts: Gut Inflammation Score (GIS), Gut Structure Score (GSS), and Gut Defense Score (GDS). A composite score was developed using the three fecal proteins: EED Protein Score (EPS).

Logistic regression models were built to test for effect modification by percent lactulose (%L) excreted, GIS, GDS, GSS, and EPS. Linear regression models were constructed to examine change in fecal host mRNA transcripts after 4 weeks of supplementation. Spearman correlation, sensitivity/ specificity analysis, and random forest classification were explored to assess whether the biomarkers measured the same domain of EED. Household WASH observations were conducted in a sub-set of study participants' homes (n=40), and summary statistics of specific behaviors were calculated by EED status (%L<0.2 vs. ≥0.2). The microbiota profile of MAM children with varying levels of EED was compared using alpha (Faith's phylogenetic diversity) and beta diversity (UniFrac distance) metrics.

Results: EED at enrollment using any of the biomarkers did not modify the effect of the study foods except for GDS (p=0.001). More children with high GDS (a sign of good gut health) recovered compared to children with lower GDS (p<0.001). There was no change in the fecal host mRNA transcripts between two time points or by study food. Among the biomarkers, only GIS and EPS were weakly but significantly correlated (r=0.22, p<0.05) with each other. None of the biomarkers predicted presence of EED as measured by lactulose:mannitol

[1] AQP9, BIRC3, CD53, CDX1, DECR1, DEFA6, HLA-DRA, IFI30, LYZ, MUC12, PIK3AP1, REG1A, REG3A, S100A8, SELL

[2] Alpha-1-Antitrypsin, Neopterin, and Myeloperoxidase

(%L≥0.2) with high sensitivity or specificity. Eight fecal host mRNA transcripts (AQP9, REG3A, IFI30, DECR1, BIRC3, SELL, PIK3AP1, DEFA6) identified EED (%L≥0.2) and severe EED (%L≥0.45) with 85% sensitivity and 80% specificity.

Differences were observed in household WASH behaviors between study participants with and without EED. These behaviors included children observed putting soil or animal feces in the mouth, animals observed drinking from household drinking water, and household drinking water storage containers having a lid. Participants with high GIS (more inflammation) had lower bacterial diversity compared to low or medium GIS (p=0.005), and different bacterial communities were present at varying levels of GIS (p=0.009).

Conclusions: MAM children who start the program with a healthier small intestine (based on GDS) are more likely to graduate from the treatment program within 12 weeks. None of the study foods performed better than the referent food in enabling MAM children with poor intestinal health to graduate from the program. Therefore, this sub-study does not support choosing any specific food for treatment of MAM based on EED status at enrollment. Most of the biomarkers examined in this sub-study measured different domains of EED. Specific WASH practices may be associated with EED, and could be targeted to address EED among MAM children.

1. BACKGROUND

Specialized nutritious foods (SNFs) have been used for many years to treat children with moderate acute malnutrition (MAM). However, the most cost-effective formulation of such foods is still to be determined, and the biological pathways by which these foods enable recovery from MAM remain poorly understood, which hampers product optimization and tailoring of complementary intervention activities. Environmental enteric dysfunction (EED), an impairment of the small intestine, might be involved in processes that lead to undernutrition, including MAM (Crane et al., 2015). EED is characterized by permeable intestinal walls, poor absorption of nutrients, and increased inflammation (Crane et al., 2015; Keusch et al., 2014). Little is known about whether EED modifies the response to SNFs during MAM treatment.

Biopsy of the small intestine is the most precise way to assess small intestinal health (Denno et al., 2014). However, this invasive method is not feasible in field conditions, nor is it appropriate for most studies. The most commonly used method to detect EED is the dual sugar or lactulose:mannitol test (L:M test). [3] In the L:M test, a mixture of lactulose and mannitol is given orally to the subject, and the ratio of the two sugars in the subject's urine describes the functional capacity of the small intestine (Denno et al., 2014). A study from Bangladesh among severely and moderately wasted children showed reduction in L:M ratio was positively associated with weight gain after nutritional intervention (Hossain et al., 2010). Despite being widely used, the dual-sugar test has drawbacks: there is a high participant burden (fasting requirements, collection of all urine excreted over four to five hours), and the test only describes two domains or characteristics of EED (permeability and absorption of the small intestine) (**Figure1**). Thus, field studies now use newer methods for characterizing EED, such as fecal host messenger ribonucleic acid (mRNA) transcripts and fecal host proteins (Arndt et al., 2016; George, et al., 2015; Kosek et al., 2013; Lin et al., 2019; Mahfuz et al., 2017; Ordiz et al., 2016).

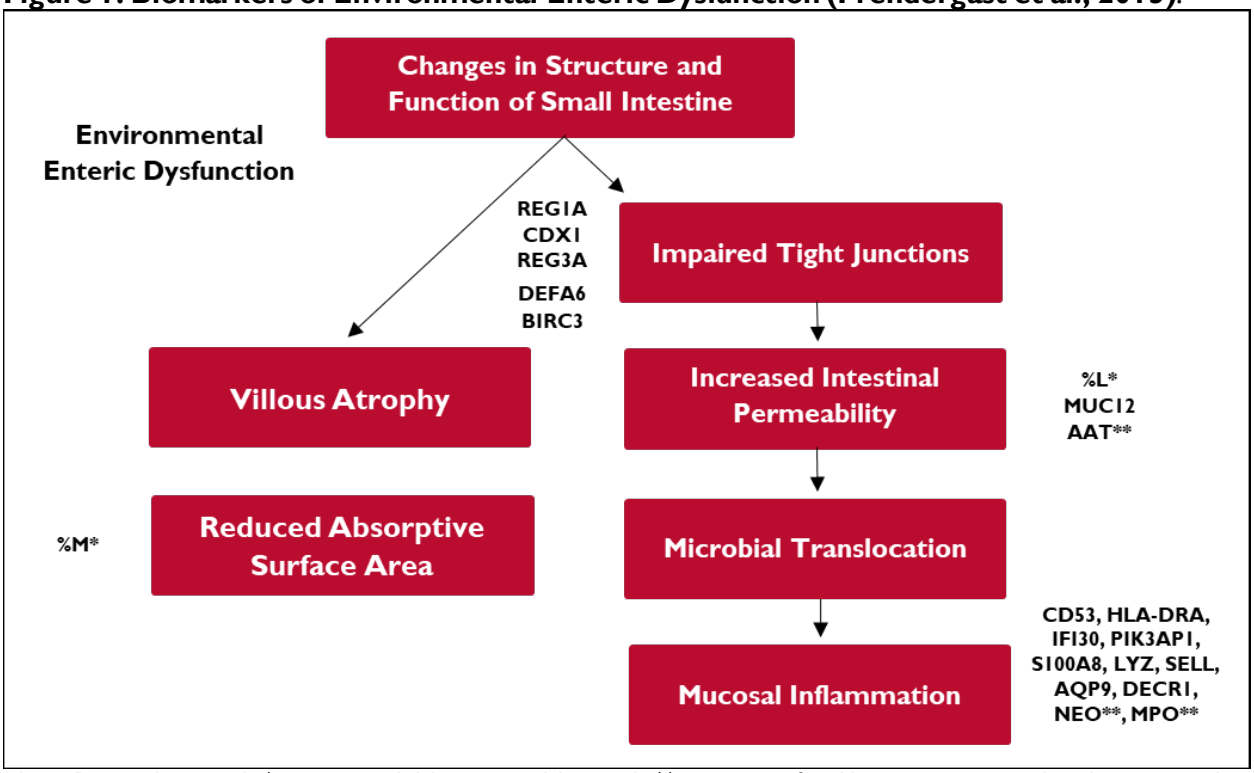
Some studies have shown that several fecal host mRNA transcripts correlate with the L:M ratio or with the percentage lactulose excreted (%L) (Agapova et al., 2013; Ordiz et al., 2016; Ordiz, et al., 2018; Yu et al., 2016). These mRNA transcripts might be more informative than the L:M test, because they capture a range of domains to measure characteristics of EED. Stool samples are also comparatively easier to collect than all urine voided over several hours. However, prediction of EED (based on L:M ratio or %L) by fecal host mRNA transcripts might be age dependent (Ordiz et al., 2016; Ordiz, et al., 2018). In this sub-study, we assessed 15 fecal host mRNA transcripts that describe three domains of EED: inflammation, permeability, and defense (Appendix Table1).[4]

The most common fecal host protein markers of EED are alpha-1-antitrypsin (AAT), neopterin (NEO), and myeloperoxidase (MPO) (Arndt et al., 2016; Campbell et al., 2017; George et al., 2015; Kosek et al., 2013; Mahfuz et al., 2017; Morita et al., 2017). The marker AAT is a measure of intestinal barrier integrity, while NEO and MPO are measures of inflammation (Kosek et al.,

[3] Lactulose is a disaccharide, and mannitol is a sugar alcohol.
[4] CDX1, HLA-DRA, MUC12, REG1A, S100A8, CD53, AQP9, BIRC3, DECR1, DEFA6, IFI30, LYZ, PIK3AP1, REG3A, SELL

2013). Although the evidence for the correlation between L:M ratio and the fecal protein markers is less clear, some studies have demonstrated a negative association between some of these markers and weight gain (Campbell et al., 2018; Kosek, 2017). A composite EED score is calculated comprising all three proteins, which we refer to as the EED Protein Score (EPS) (Kosek et al., 2013).

Figure 1. Biomarkers of Environmental Enteric Dysfunction (Prendergast et al., 2015).



Note: Biomarkers with * represent L:M test variables, with ** represent fecal host proteins, and with no asterisks represent fecal host mRNA transcripts.

While there is a need to identify field-appropriate biomarkers of EED, it is equally important to understand the factors that might contribute to EED. Observational studies have demonstrated positive associations between EED and different sub-optimal water, sanitation, and hygiene (WASH) practices, including water quality, sanitation, handwashing, animals corralled in a room where children sleep, and geophagy (soil consumption) (George et al., 2015; Lauer et al., 2018; Lin et al., 2013). Additionally, there is growing recognition that alterations of the host microbiota (microorganisms that colonize the gut) play a critical role in the health and nutrition status of young children. Studies have found differences in the microbiota of malnourished children compared to their healthy peers (Subramanian et al., 2014). There is also some evidence for variation in the prevalence of certain types of bacteria among children with low and high EED (Ordiz et al., 2017). These findings suggest that the microbiota profile of a malnourished child might indicate the level of EED detected.

The objectives of the EED sub-study were, therefore, to examine 1) whether EED modifies the effects of SNFs on recovery from MAM (mid-upper arm circumference [MUAC] greater than

or equal to 12.5cm), 2) if there is an improvement in EED over four weeks of treatment with SNFs, 3) how EED biomarkers compare with each other and with measures of intestinal inflammation and damage, 4) the association between WASH practices and EED, and 5) the association between the microbiota profile of MAM children and their varying levels of EED. Results from the sub-study will provide information on the role of EED in children with MAM and the ability of supplementary foods to achieve growth in the presence of EED, which could be used to guide selection of SNFs in future MAM treatment programs.

2. METHODS

2.1 Study Design

The EED sub-study was nested within the *Four Foods MAM Treatment Study*. The latter was a prospective, cluster-randomized, controlled clinical and cost-effectiveness trial assessing four SNFs to treat children age 6-59 months with MAM, defined as MUAC greater than or equal to 11.5 cm and less than 12.5 cm without bipedal edema. While the main study was conducted in 29 peripheral health units (PHUs) randomly assigned to deliver one of four isoenergetic foods, the EED sub-study was conducted at eight of the study PHUs, two per arm, purposively selected based on logistical constraints. Biological samples were collected from July 2017 to August 2018 in collaboration with the Project Peanut Butter (PPB) study staff. The PHUs where the EED sub-study was conducted were:

- Corn Soy Blend Plus (CSB+) with oil: Gissiwolu and Nyandehun
- Super Cereal Plus with Amylase (SC+A): Gbongay and Wai
- Corn Soy Whey Blend (CSWB) with oil: Bandasuma and Gbaa
- Ready-to-Use Supplementary Food (RUSF): Jendema and Sengama

2.2 Approvals

The sub-study was approved by the Sierra Leone Ethics and Scientific Review Committee and the Tufts University Health Sciences Institutional Review Board. Written informed consent was obtained from caregivers of all sub-study participants.

2.3 Participants and Sample Size

To examine the effects of the study foods on recovery from MAM in the presence of EED, the total planned sample size was 404 (101 per arm). This achieves 80 percent power to detect an R-squared value of 0.2 in a multivariable regression model with eight predictors at a significance level (alpha) of 0.05, assuming a design effect of 1.2. To examine if there is an improvement in EED due to supplementary feeding, the planned sample size was 230 (approximately 58 per arm). This achieves 80 percent power to detect a mean of paired differences of 0.153 copies of the mRNA transcripts with an estimated standard deviation of differences of 0.846 and with a significance level (alpha) of 0.05 using a two-sided paired t-test, assuming a design effect of 1.2. The sample size for WASH observations, fecal host protein markers, and microbiota analysis was based on logistical and budgetary constraints. Participants enrolled at the EED sub-study sites over a three-month period were invited to participate in the WASH observations. For the fecal host protein marker analysis, the sample size was capped at 200 (approximately 50 per

arm). For the microbiota analysis, the sample size was capped at 100: 25 for the three levels of EED (low, medium, high) independent of arm, and 25 healthy (non-MAM) controls.

2.4 EED Sample Collection Methodology

At the eight EED sub-study sites, caregivers of eligible children who consented to the *Four Foods MAM Treatment Study* were invited to participate in the sub-study. In this sub-study, EED was measured using three biomarkers: L:M test, fecal host mRNA transcripts, and fecal host proteins **(Table 1).**

Table 1. EED Sub-Study Biomarker Sample Type and Collection Time Point

Biomarker	Sample type	Time point
L:M test	Urine	Enrollment
Fecal host mRNA transcripts	Stool	Enrollment and after 4 weeks
Fecal host proteins	Stool	Enrollment

[1] L:M test: Because it was known from the beginning that the study would also be using mRNA transcripts and proteins, and because these methods provide more detailed information than the L:M test, the research team decided it was in the best interests of the children to expose them to this method only once, at baseline, to reduce respondent burden. The L:M test was administered as follows At the health facility, a 20 ml dose containing 5 g of lactulose (McKesson, San Francisco, CA) and 1 g of mannitol (Sigma Aldrich, St. Louis, MO) was given to each participant orally, using a medicine cup or syringe, after an eight-hour overnight fast. Water was allowed as often as desired by the child prior to and after being dosed with the sugar solution. After dosing, a urine collection bag (Thermo Fisher Scientific, Waltham, MA) and a locally prepared, non-absorbent diaper were attached to the participant.

All urine excreted over the next four hours was collected in a cup with 10 mg of thimerosal (Sigma Aldrich, St. Louis, MO) to prevent bacterial degradation of the sugars. The urine was mixed with a pipette, transferred to plastic cryovials, and flash frozen in liquid nitrogen at the PHU. Participants and caregivers were provided with lunch after the three-hour mark, when breastfeeding was also allowed. The total urine volume was recorded using a graduated cylinder. Every month, samples were transferred to a -20°C freezer at the University of Makeni prior to being shipped on dry ice to Baylor College of Medicine (TX). The concentration of the sugars in the urine was analyzed by high-performance liquid chromatography (Shulman et al., 1998).

Table 2. Calculation of L:M Test Variables

Variable		Calculation
%Lactulose excreted (%L)	=	$\frac{\text{Concentration of lactulose in urine sample * 4 hour volume of urine}}{\text{Concentration of lactulose in the dose}}$
%Mannitol excreted (%M)	=	$\frac{\text{Concentration of mannitol in urine sample * 4 hour volume of urine}}{\text{Concentration of mannitol in the dose}}$
Lactulose: Mannitol Excretion Ratio (LMER)	=	$\frac{\text{\%Lactulose excreted (\%L)}}{\text{\%Mannitol excreted (\%M)}}$
L: M ratio	=	$\frac{\text{Concentration of lactulose in urine sample}}{\text{Concentration of mannitol in urine sample}}$

A concentrate of the two sugars in the samples collected was used to calculate the L:M test variables **(Table 2)**. Although the lactulose:mannitol excretion ratio (LMER) or L:M ratio have historically been the markers of EED, new evidence suggests that both high or low LMER or L:M ratio could represent poor gut function and that %L might be a more accurate measure of intestinal health (Ordiz, Davitt, et al., 2018). For this reason, our primary L:M test indicator is %L, but we also conducted the analysis with LMER and L:M ratio. Severity terciles of %L were generated based on existing literature **(Table 3)** (Ordiz et al., 2016). Presence of EED was determined as values of %L≥0.2, which included both medium and high terciles (Yu et al., 2016).

[2] Fecal markers: Stool samples were collected at any point prior to, during, or after the four-hour wait period for the L:M test. Once a participant had a bowel movement, the diaper was removed, and all stool collected was mixed with a spatula and transferred into plastic cryovials without any fixative. The cryovials with stool were flash frozen in liquid nitrogen at the PHU. Every month, samples were transferred to a -80°C freezer at the University of Makeni prior to being shipped on dry ice to labs for analysis. Fifteen fecal host mRNA transcript were analyzed by digital droplet polymerase chain reaction at Washington University School of Medicine (St. Louis, MO). Fecal host proteins were analyzed by Enzyme Linked Immunosorbent Assay (kits from R&D Systems for AAT and MPO, and GenWay Biotech for NEO) at the USDA Human Nutrition Research Center on Aging at Tufts University (MA) (Agapova et al., 2013; Kosek et al., 2013; Stauber et al., 2016). Microbiota analysis was conducted by 16S ribosomal RNA sequencing at the Tufts Medical Center (MA) (Caporaso et al., 2011).

The fecal host mRNA transcript concentrations were continuous variables denoting copies of the mRNA transcripts per copy of GAPDH[5], the transcript to which all other transcripts were normalized. Using these concentrations, the fecal host mRNA transcript EED scores were developed via factor analysis. Three fecal host mRNA transcript EED scores that resulted from

[5] Glyceraldehyde 3-phosphate dehydrogenase (GAPDH)

the factor analysis are referred to here as Gut Inflammation Score (GIS), Gut Structure Score (GSS), and Gut Defense Score (GDS), based on functions of the mRNA transcripts that grouped together[6]. The GIS tercile was constructed by dividing the continuous GIS variable into three equal groups to examine severity using this biomarker **(Table 3)**. The process was followed for GSS and GDS. We speculate that high GIS and GSS, but low GDS are indicative of EED.

The three fecal host protein concentrations were continuous parameters. To construct the EPS, we first calculated the 25th and 75th percentile for each of the three proteins separately. Then a new variable was generated for each protein, which took the value 0 for <25th percentile, 1 for 25th-75th, and 2 for >75th percentile. These values were input in the formula EPS = 2(AAT category) + 2(MPO category) + 1(NEO category), and the resulting score ranged from 0 to 10 (Kosek et al., 2013). The EPS terciles were constructed by dividing the continuous EPS variable into three equal groups to examine severity using this biomarker **(Table 3)**:

Table 3. Severity Cutoffs for EED Biomarkers

Biomarker	Low	Medium	High
%L	<0.2	0.2-0.44	≥0.45
GIS	T1	T2	T3
GSS	T1	T2	T3
GDS	T1	T2	T3
EPS	T1	T2	T3

Note: T1, First tercile; T2, Second tercile; T3, Third tercile; GIS, Gut Inflammation Score; GSS, Gut Structure Score; GDS, Gut Defense score; EPS, EED Protein Score

[3] WASH observations: All caregivers of participants who consented to the EED sub-study from June to August 2018 were invited to participate in the WASH observations (n=70). Seven members of the Sierra Leone Red Cross Society, Pujehun branch, were trained for five days on consent and the observation procedures, including one full six-hour practical training. These staff conducted direct observation of the study child and the surrounding environment for six hours in each household, from 7:00 a.m. to 1:00 p.m. The Red Cross study staff used a paper-based semiquantitative form to record their observations. This form was divided into four sections: 1) mouthing (placing something in the mouth, but not eating), 2) eating, 3) defecation/urination, and 4) spot check (assessment for hand cleanliness of child and caregiver, and status of latrine, drinking water, compound, and animals).

2.5 Data management: PPB study staff recorded data for biological samples on paper clinic cards and then entered data from the cards into an electronic database (KoBoCollect) at the PHU. Tufts research assistants cross-checked electronic data entry against the clinic cards on a

[6] The mRNA transcripts AQP9, CD53, IFI30, PIK3AP1, S100A8, and SELL correlated highly with GIS; BIRC3, CDX1, AND MUC12 correlated highly with GSS, and DEFA6 and REG3A correlated highly with GDS. We hypothesize that high GIS and GSS but low GDS represent EED.

monthly basis. One research assistant entered data for the WASH observations into an electronic database (KoBoCollect) after completion of observations on all participants.

2.6 Hypotheses

1. **The presence of EED at enrollment modifies the effect of the four study foods on recovery (MUAC≥12.5cm) from MAM within 12 weeks.**
   Measurements: The outcome was the binary variable recovery from MAM within 12 weeks. The EED variables were terciles for %L, GIS, GSS, GDS, and EPS.

   a) One of the four study foods achieves recovery better than another in the presence of EED.
   Measurements: The outcome was the binary variable recovery from MAM within 12 weeks. The exposure was a categorical variable for study food. The effect modifiers were the continuous EED variables: %L, GIS, GSS, GDS, and EPS in separate models. An interaction term for study food and the EED indicator was included in each model to assess effect modification.

2. **There is an improvement in EED after four weeks of treatment.**
   Measurement: The continuous variable for concentration for each mRNA transcript at enrollment and after four weeks of treatment were compared.

   a) One of the four foods performs better than another at improving EED after four weeks of treatment.
   Measurements: The outcome was the continuous variable change in mRNA transcript concentration from enrollment to after four weeks. The exposure was a categorical variable for study food, and the covariate was the continuous variable for the mRNA transcript concentration at enrollment. Separate models were run for each transcript.

3. **The EED biomarkers are correlated with each other and can correctly identify MAM children with %L≥0.2.**
   Measurements: The continuous EED variables were %L, GIS, GSS, GDS, EPS, and 15 fecal host mRNA transcripts. The categorical EED variables were %L (binary), and two binary variables each for GIS, GSS, GDS, and EPS constructed by regrouping the respective terciles.

4. **Household WASH practices and microbiota profile are associated with EED at baseline among MAM children 6-59 months of age.**
   Measurements: For WASH, the outcome was the binary variable for EED, %L. The WASH variables were as follows: children putting soil/animal feces in mouth, animals drinking from household drinking water, clean compound, clean caregiver hands, dirt floor of dwelling, no access to latrine, and household drinking water storage container has lid. For microbiota analysis, the diversity metrics were alpha diversity (Faith's phylogenetic diversity) and beta diversity (unweighted UniFrac distance). The diversity

metrics were examined for categorical variables of EED: %L tercile, GIS tercile, GSS tercile, GDS tercile, and MAM status.

Anthropometry and covariate data were extracted from the main dataset of the *Four Foods MAM Treatment Study* for analysis of the above-mentioned hypotheses.

3. Statistical Analysis

A total of 601 individuals were enrolled in the EED sub-study, with urine samples collected from 422 participants and stool samples collected from 475 participants at enrollment **(Figure 2)**. Urine samples were not collected from 179 participants due to illness or no show. Stool samples were not collected from 126 participants due to illness, no show, or no bowel movement at PHU. All defaulters[7] were excluded from statistical analysis (n=34 for urine and n=36 for stool). Twenty-nine stool samples were excluded from analysis because of low levels of GAPDH in the lab reports. Stool samples were also collected from 277 subjects after four weeks of treatment. Stool samples were not collected from 198 participants who provided a sample at enrollment because of no show, illness, transfer to severe acute malnutrition (SAM), death, or they only consented to sample collection at one time point.[8] Four stool samples were excluded because of low levels of GAPDH as reported by the lab.

Fecal protein markers were examined in 190 stool samples collected at enrollment. Microbiota analysis was conducted on 78 samples. Among the healthy participants, 32 caregivers were recruited, and samples were collected from 21 participants. Samples were not collected from 11 healthy participants because of no bowel movement at the PHU.

Statistical analyses were conducted using Stata 15 software (StataCorp, College Station, TX). First, chi-square test was used to assess whether any of the EED biomarker terciles at baseline predicted recovery (binary) from MAM irrespective of study food. Then, unadjusted and adjusted logistic regression was used to examine modification of the effect of the study foods (categorical) on recovery from MAM (binary) by EED (continuous). To do this, we first ran the model with EED and study food as predictors. Then an interaction term for the EED variable and study food was included in the model. Separate models were run for each EED biomarker. Covariates for the adjusted model were selected based on biological relevance (age, gender, and previous SAM status of the child).

Paired t-test was used to assess differences in each mRNA transcript concentration (continuous) from enrollment to after four weeks of treatment irrespective of study food. Then, unadjusted and adjusted linear regression models were used to determine whether there was a difference in change in each mRNA transcript (continuous) from enrollment to after four weeks of treatment by study food (categorical). The adjusted model included age, gender, previous SAM status, and baseline concentration of the transcript. The continuous mRNA

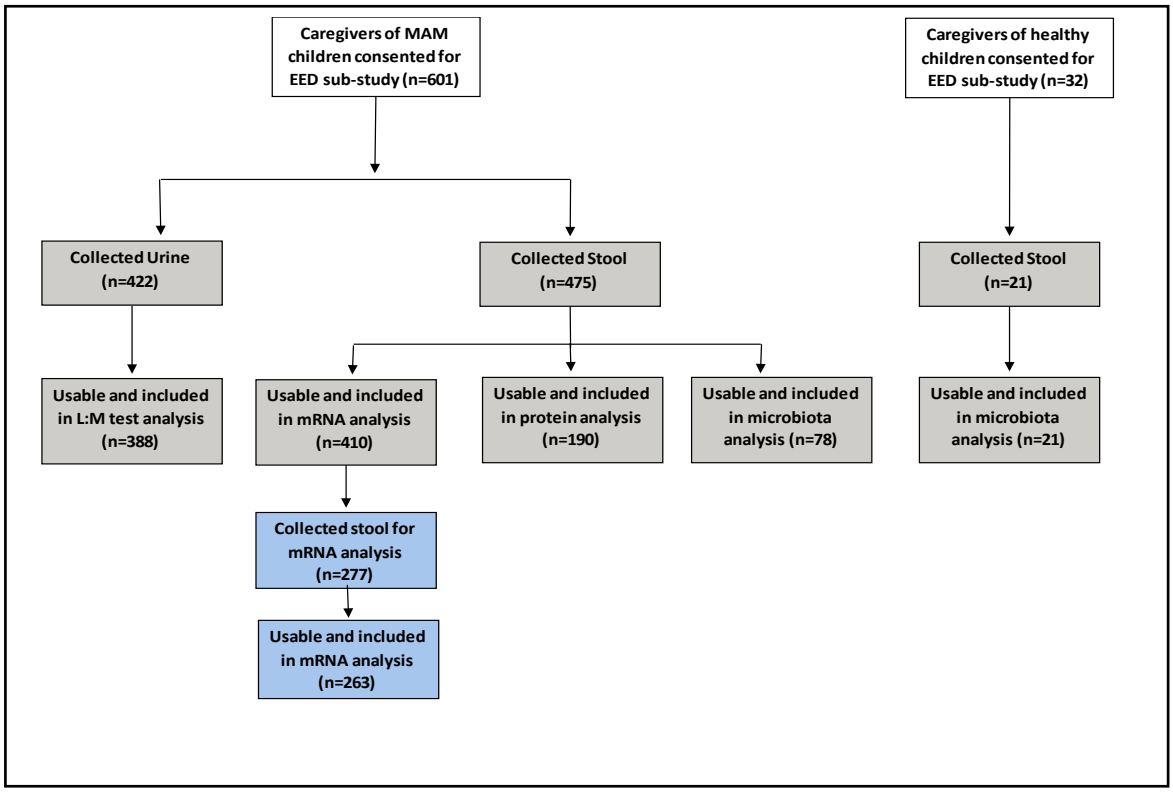
[7] Participants who missed three consecutive visits
[8] Samples were collected from some participants only at enrollment because the sub-study had a definite end date that would not allow four-week sample collection for these participants.

transcripts were natural log transformed because they were not normally distributed. Statistical significance was set at p-value <0.05 for all analyses.

Since the sub-study used three biomarkers to examine EED, tests were run to determine whether the biomarkers described the same domains of EED (permeability, inflammation, defense). This was done by assessing the Spearman correlation among the biomarkers, and the ability of the mRNA scores and EPS to differentiate participants with and without EED based on the dichotomous variable for %L (<0.2 vs. ≥0.2). The standard of %L was used because it is the measure with the strongest physiologic basis as a direct measure of gut integrity. Additionally, the random forest classification model was estimated using the Stata "randomforest" module to assess how well the 15 fecal host mRNA transcripts could predict %L≥0.2 and %L≥0.45. This approach was also examined because we suspect that the relationship between the mRNA transcripts and %L might not be linear.

Summary statistics were calculated for the variables "putting soil/animal feces in mouth," "animals drinking from household drinking water," "clean compound," "clean caregiver hands," "dirt floor of dwelling," "no access to latrine," and "household drinking water storage container has lid." Differences in these characteristics were presented for participants stratified by the dichotomous variable for %L (<0.2 vs. ≥0.2) (n=40).

Figure 2. Samples Collected for EED Sub-Study Analysis at Enrollment and After Four Weeks



Note: Boxes in gray are samples collected at enrollment and boxes in blue are samples collected after four weeks. Out of the 601 participants who consented, 70 were included in the WASH observations.

For the microbiota profile, statistical analysis was conducted using QIIME2 version 2018.8 (http://www.qiime2.org) on 78 MAM participants with varying levels of EED and on 21 healthy participants. The differences in bacterial diversity were examined at varying levels of EED (using different EED biomarkers), and between children with and without MAM. The diversity metrics were Faith's phylogenetic diversity for alpha diversity and unweighted UniFrac distance for beta diversity. Alpha diversity measures the within-group diversity: the number of bacterial species within a group (Lozupone & Knight, 2008). The alpha diversity between groups can be tested using the Kruskal-Wallis test. A high alpha diversity score is generally considered a sign of better microbiota profile. Beta diversity measures the between-group diversity: bacterial species that are dissimilar between two groups (Lozupone & Knight, 2008). A high beta diversity represents more dissimilarities between groups. Beta diversity is first examined visually through principal coordinates analysis (PCoA) plots, and then permutational multivariate analysis of variance (PERMANOVA) can be used to statistically test differences in centroids between groups.

4. Results

4.1 Study Population

Background characteristics of study participants that contributed to at least one EED biomarker result were balanced across the arms except for gender and previous SAM status **(Table 4)**. The participants' mean age ± standard deviation (SD) was 13.96 ± 8.71 months, 58 percent were female, and 23 percent had transferred from SAM. Caregivers reported that 77 percent of participants were currently breastfeeding, 6 percent had experienced diarrhea in the past seven days, and 54 percent experienced severe household food insecurity. The anthropometric measurements were balanced across arms. Overall, 68 percent of sub-study participants graduated from the treatment program.

4.2 Prevalence of EED

The median (25^{th}, 75^{th} percentile) of %L at enrollment was 0.34 (0.21, 0.73) **(Table 5)**; 77 percent of participants had EED (%L≥0.2) **(Figure 3)**. The median and interquartile range of MPO was 42,173 (18,895, 88,332) ng/mL, and NEO was 940 (456, 1,874) nmol/L. These values were higher than non-tropical standards of ≤2,000 ng/mL for MPO and <70 nmol/L for NEO (Beckmann G, 2000; Ledjeff, Artner-Dworzak, Witasek, Fuchs, & Hausen, 2001; Saiki, 1998). Using these standards, 98 percent of sub-study participants had elevated MPO and 99 percent had elevated NEO. The median and interquartile range for AAT was 2,217 (1,756, 2,916) ng/mL, which was lower than the non-tropical standard of <0.27 mg/g. Using this standard, none of the study participants had elevated AAT.

Figure 3. Prevalence of EED (%L≥0.2)

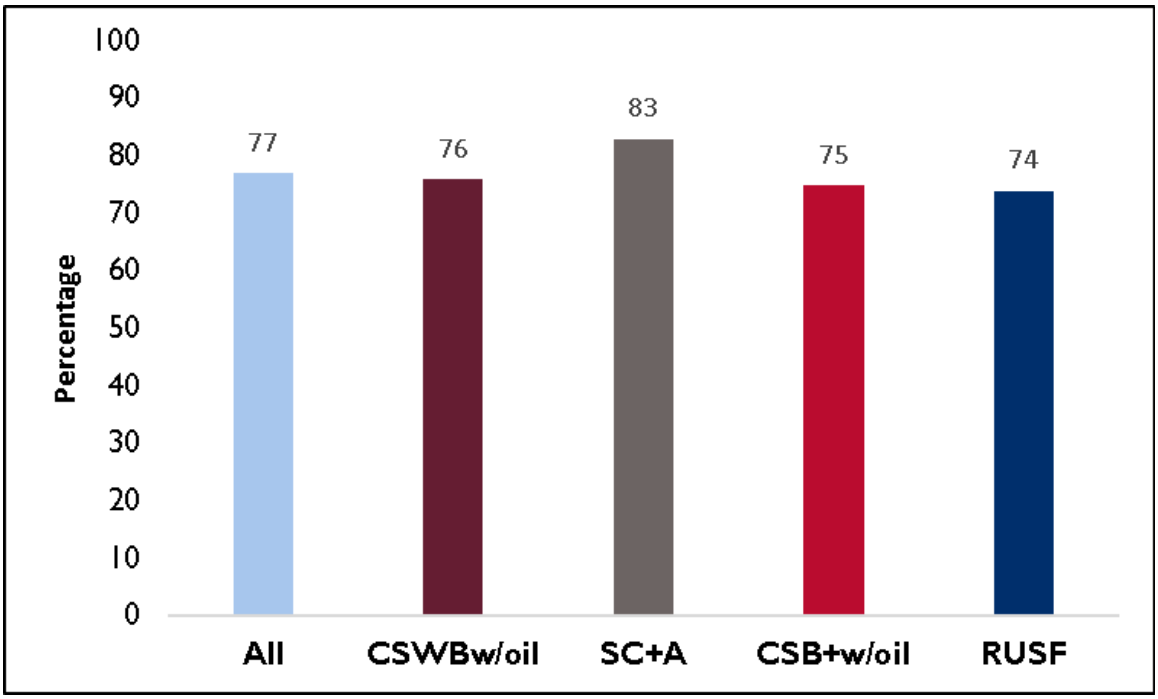


Table 4. Characteristics of EED Sub-Study Participants[§]

	All	CSWB w/oil	SC+A	CSB+ w/ oil	RUSF	P-value[¶]
n	484	74	142	122	146	
Enrollment						
Age (months)	13.96±8.71	12.67±8.65	14.15±9.13	13.43±7.66	14.86±9.12	0.299
Female	283(58)	46(62)	94(66)	56(46)	87(60)	0.008*
Transferred from SAM	112(23)	16(22)	34(24)	39(32)	23(16)	0.019*
Currently breastfed	368(77)	61(82)	106(76)	96(79)	105(72)	0.335
Diarrhea in past 7 days	31(6)	5(7)	10(7)	7(6)	9(6)	0.978
HFIAS						0.070
Food secure	160(33)	29(39)	38(27)	43(35)	50(34)	
Moderately food insec	64(13)	15(20)	17(12)	11(9)	21(14)	
Severely food insecur	260(54)	30(41)	87(61)	68(56)	75(51)	
Anthropometry						
MUAC	11.97±0.27	12±0.27	11.95±0.26	11.94±0.27	11.99±0.27	0.311
LAZ	-2.78±1.23	-2.65±1.22	-2.75±1.24	-2.88±1.25	-2.77±1.2	0.646
WLZ	-1.82±0.76	-1.66±0.8	-1.77±0.73	-1.84±0.79	-1.93±0.72	0.064
WAZ	-2.93±0.84	-2.86±0.87	-2.85±0.81	-3±0.88	-2.99±0.8	0.344
Outcome						
Graduated	327(68)	48(65)	93(65)	78(64)	108(74)	0.262

[§]Cells represent Mean±SD or n(%)

*p<0.05

[¶]ANOVA for continuous variables and chi-square test for categorical variables

Abbreviations: SAM, severe acute malnutrition; HFIAS, household food insecurity access scale; MUAC, mid-upper arm circumference; LAZ, length-for-age z score; WLZ, weight-for-length z score; WAZ, weight-for-age z score

Hypothesis: The presence of EED at enrollment modifies the effect of the four study foods on recovery (MUAC≥12.5cm) from MAM within 12 weeks.

Recovery from MAM did not vary by level of EED at enrollment as indicated by %L tercile, EPS tercile, GIS tercile, or GSS tercile. However, there was a statistically significant difference by level of the GDS tercile **(Figure 4)**. A significantly larger proportion of participants from the high group of the GDS tercile at enrollment graduated from the program compared to participants from the medium or low group using the chi-square test (P<0.001).

We also examined the above-mentioned EED biomarkers as modifiers of the effect of the study foods on recovery from MAM within 12 weeks. Since the model with the terciles was not stable when performing diagnostic tests, we used the continuous forms of the variables. Of the five biomarkers, only GDS was a significant modifier in both unadjusted model (P=0.002) and model adjusted for age, gender, and previous SAM status (P=0.001) **(Figure 5)**. With the exception of RUSF, we found no statistically significant differences between the study food and the comparator CSB+ with oil when controlling for multiple comparisons using the Bonferroni method; RUSF performed better than CSB+ with oil at high levels of GDS (P=0.035), which we suspect is a sign of good gut health. We did not find effect modification by LMER or L:M ratio.

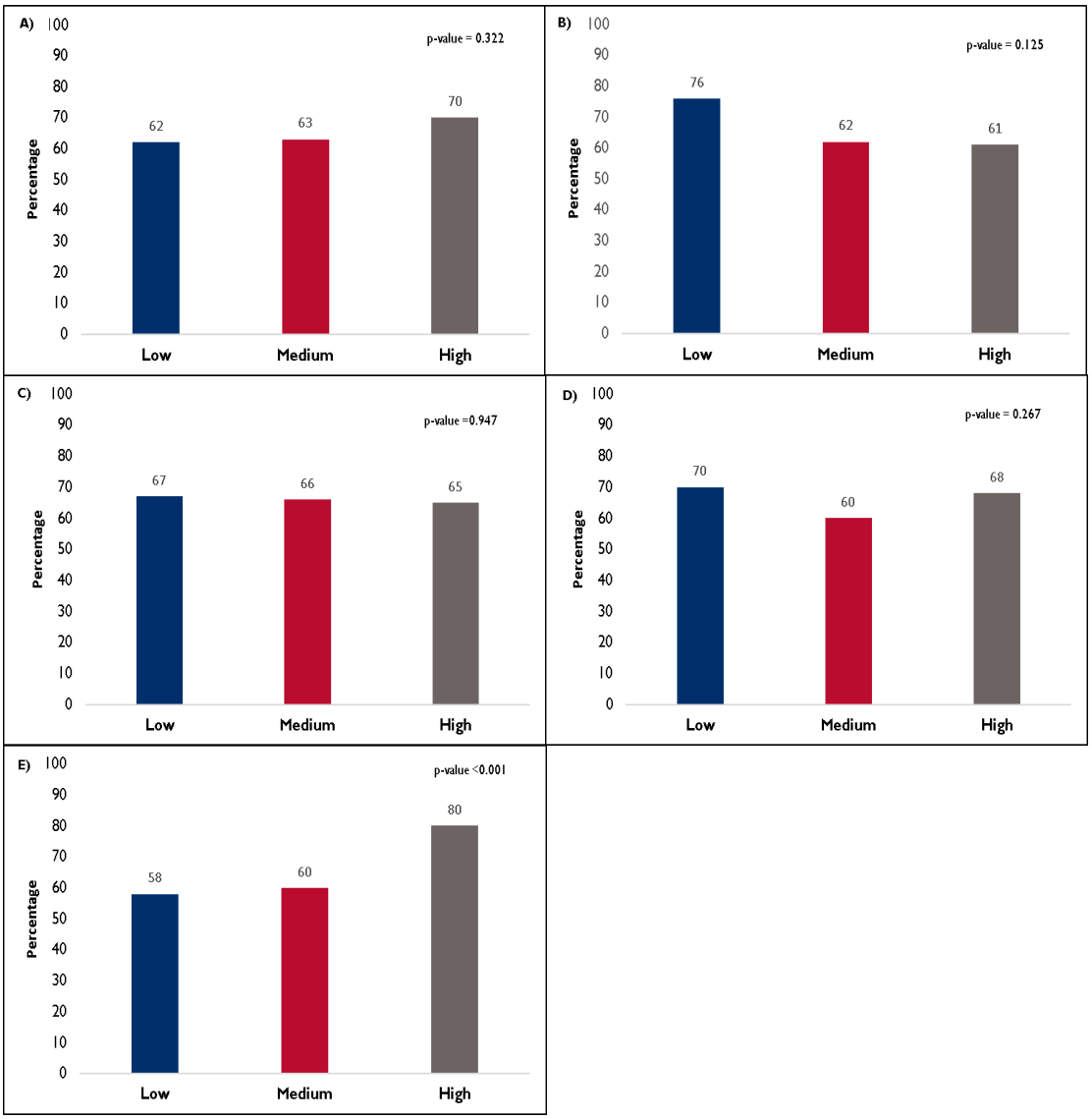
Table 5. EED Biomarkers of Sub-Study Participants at Enrollment

	n¶	Median (25th, 75th percentile)
LM test		
LMER	388	0.10 (0.06, 0.15)
LM Ratio	388	0.48 (0.32, 0.73)
%Lactulose	388	0.34 (0.21, 0.62)
%Mannitol	388	3.86 (2.42, 5.65)
mRNA§		
AQP9	398	0.13 (0.06, 0.31)
BIRC3	399	0.21 (0.11, 0.37)
CD53	405	0.20 (0.07, 0.55)
CDX1	406	0.05 (0.03, 0.08)
DECR1	400	0.06 (0.03, 0.09)
DEFA6	400	0.08 (0.04, 0.21)
HLA-DRA	404	0.14 (0.07, 0.25)
IFI30	400	0.31 (0.14, 0.64)
LYZ	410	0.11 (0.05, 0.21)
MUC12	406	0.43 (0.23, 0.84)
PIK3AP1	400	0.16 (0.06, 0.38)
REG1A	406	0.15 (0.06, 0.36)
REG3A	400	0.07 (0.03, 0.15)
S100A8	403	1.19 (0.53, 2.83)
SELL	400	0.06 (0.02, 0.16)
TNF	124	0.06 (0.01, 0.02)
Protein		
AAT (ng/mL)	190	2,217.49 (1,756.40, 2,915.82)
MPO (ng/mL)	190	42,172.51 (18,895.35, 88,332.42)
NEO (nmol/L)	190	939.71 (456.18, 1,873.8)
EED Score	190	5 (4, 6)

§Copies per copy of GAPDH

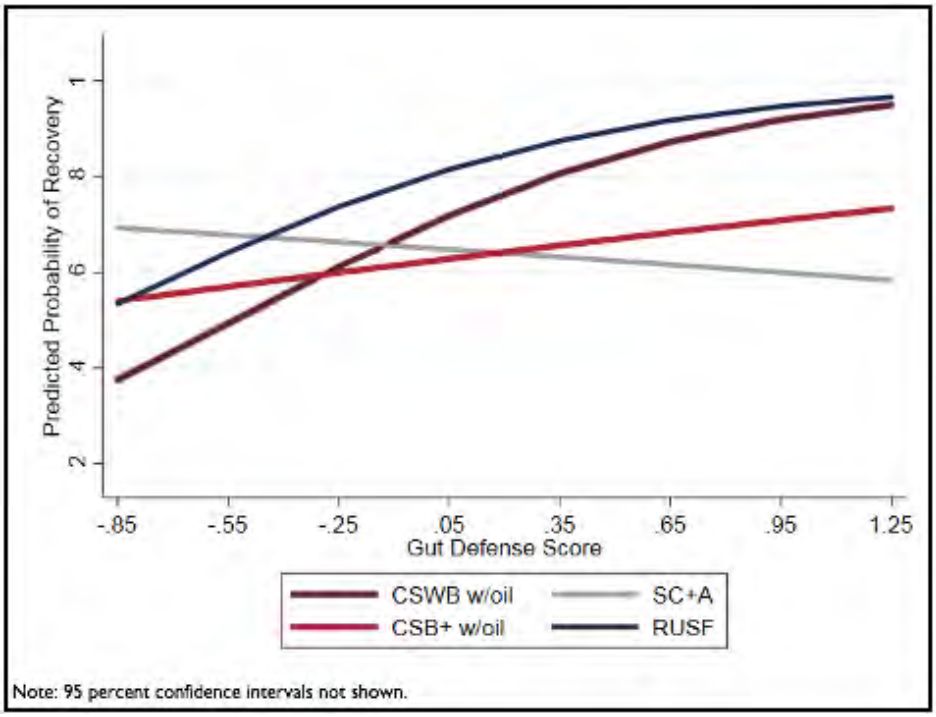
¶One outlier for LYZ excluded

Figure 4. Recovery from MAM by Level of EED at Enrollment



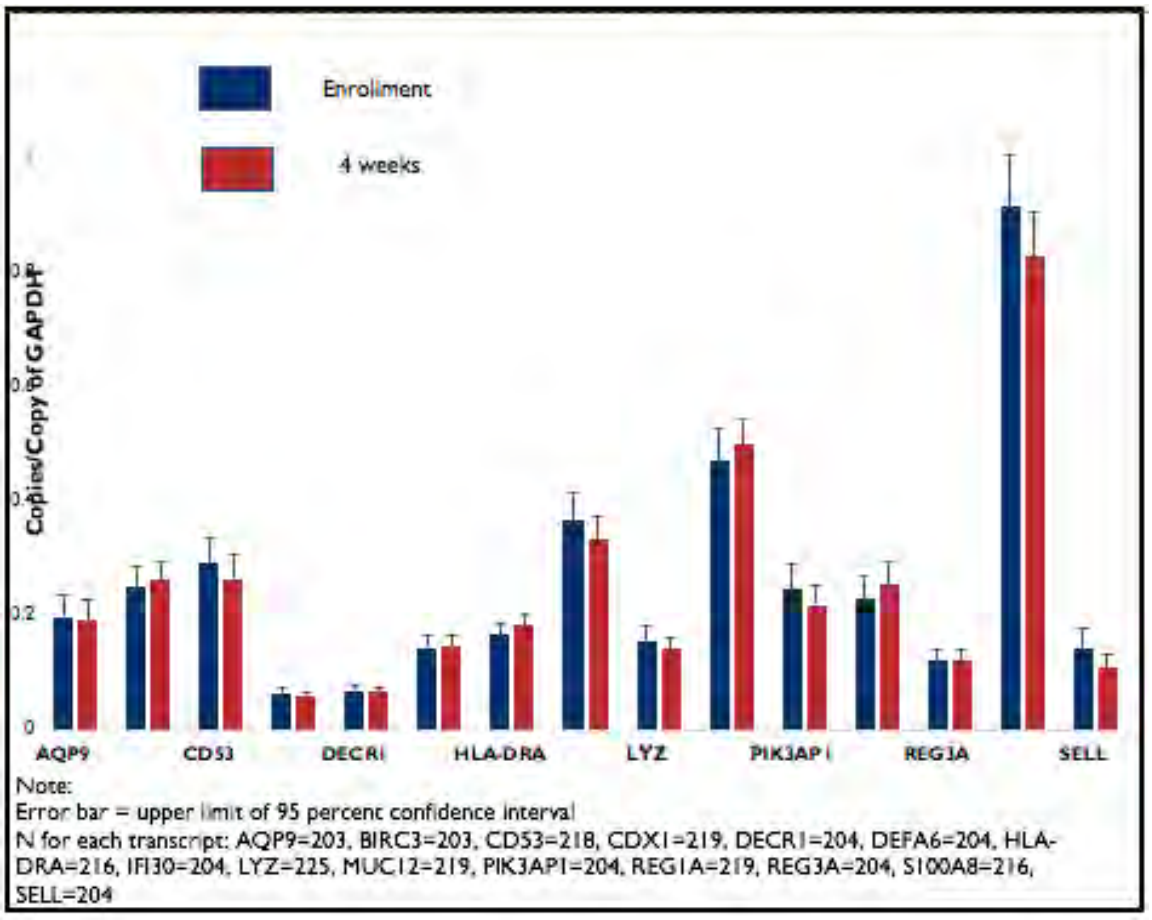
Note: P-values are from Chi-square test: A) %L, B) EED Protein Score, C) Gut Inflammation Score, D) Gut Structure Score, E) Gut Defense Score.

Figure 5: Predicted Probability of Recovery by Study Food and GDS at Enrollment.



Hypothesis: There is an improvement in EED after four weeks of treatment.
The concentration of 15 fecal host mRNA transcripts changed slightly between enrollment and after four weeks on the study foods **(Figure 6)**. However, these changes were not large enough to detect a statistically significant difference. We also examined whether this difference varied by study food and did not find a statistically significant difference for any of the mRNA transcripts.

Figure 6. Mean mRNA Transcript Concentration at Enrollment and After Four Weeks on the Study Foods



Hypothesis: The EED biomarkers are correlated with each other and can correctly identify MAM children with %L≥0.2.

The correlation among %L, GIS, GSS, GDS, and EPS was weak **(Table 6)**. However, there was a significant correlation between GIS and EPS ($r=0.22$, $p<0.05$) and a significant but negative correlation between GIS and GSS ($r=-0.19$, $p<0.05$). The significant positive correlation between GIS and EPS suggests that both are measuring the same characteristics of EED. Since EPS is a marker of inflammation, we can infer that GIS also detects inflammation during EED. Given that the L:M test is the most widely used EED biomarker, assessing the ability of newer biomarkers to differentiate participants with and without EED based on the L:M test is of interest. We find that GIS, GSS, GDS, and EPS poorly differentiated participants with and without EED based on %L as demonstrated by the fact that none of the biomarkers could predict the presence of EED (%L≥0.2) with high sensitivity and specificity (>0.8) **(Table 7)**. These results suggest that the EED biomarkers used in this sub-study are measuring different aspects or stages of EED, as was suspected.

Using the random forest classification method, we found that eight fecal host mRNA transcripts[9] were important predictors of %L≥0.2 **(Figure 7)**. Furthermore, a model with these fecal host mRNA transcripts was able to identify %L≥0.2 with 100 percent sensitivity and 80 percent specificity. These same fecal host mRNA transcripts were also able to identify %L≥0.45 (severe EED) with 85 percent sensitivity and 80 percent specificity. Unlike the factor analysis-based scores that separately identify a particular domain of EED, these eight transcripts together likely capture many domains of EED.[10]

Table 6. Spearmen Correlation Among Untransformed EED Biomarkers

	%L[§]	GIS	GSS	GDS	EPS
%L	1				
GIS	-0.06	1			
GSS	-0.00	-0.19*	1		
GDS	-0.05	0.02	0.02	1	
EPS	0.08	0.22*	-0.13	-0.14	1

*P<0.05

[§]%L, Percentage Lactulose Excreted; GIS, Gut Inflammation Score; GSS, Gut Structure Score; GDS, Gut Defense Score; EPS, EED Protein Score

Table 7. Sensitivity and Specificity of EED Scores for %L (<0.2 Vs. ≥0.2)

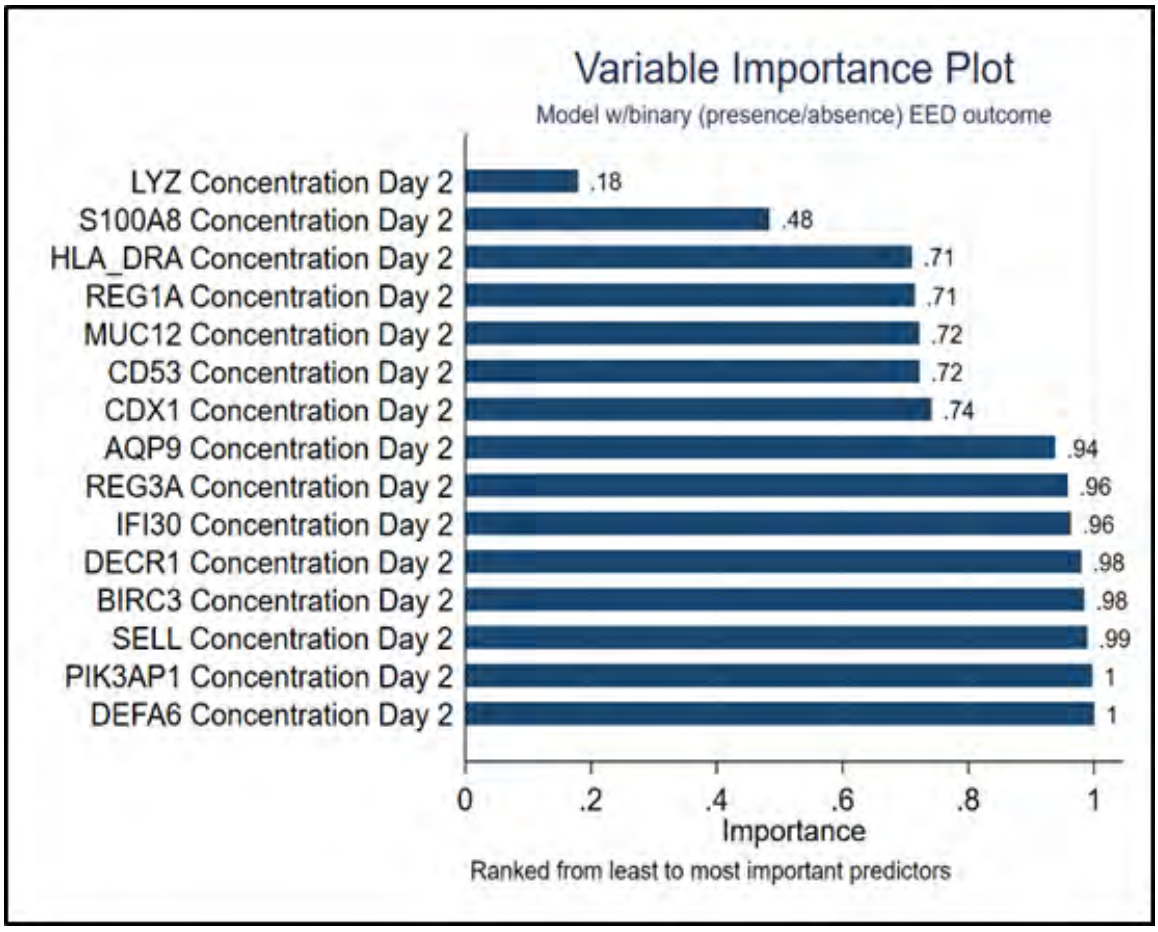
	Low vs. Medium & High		Low & Medium vs. High	
	Sensitivity	Specificity	Sensitivity	Specificity
GIS[§]	66%	34%	31%	61%
GSS	63%	25%	33%	61%
GDS	66%	36%	34%	69%
EPS	62%	39%	27%	78%

[§]GIS, Gut Inflammation Score; GSS, Gut Structure Score; GDS, Gut Defense Score; EPS, EED Protein Score

[9] AQP9, REG3A, IFI30, DECR1, BIRC3, SELL, PIK3AP1, and DEFA6

[10] AQP9, IFI30, DECR1, SELL, and PIK3AP1 – inflammation; BIRC3 –structure; and REG3A and DEFA6 – defensins (anti-microbials)

Figure 7. Variable Importance Plots from Random Forest Classification Models

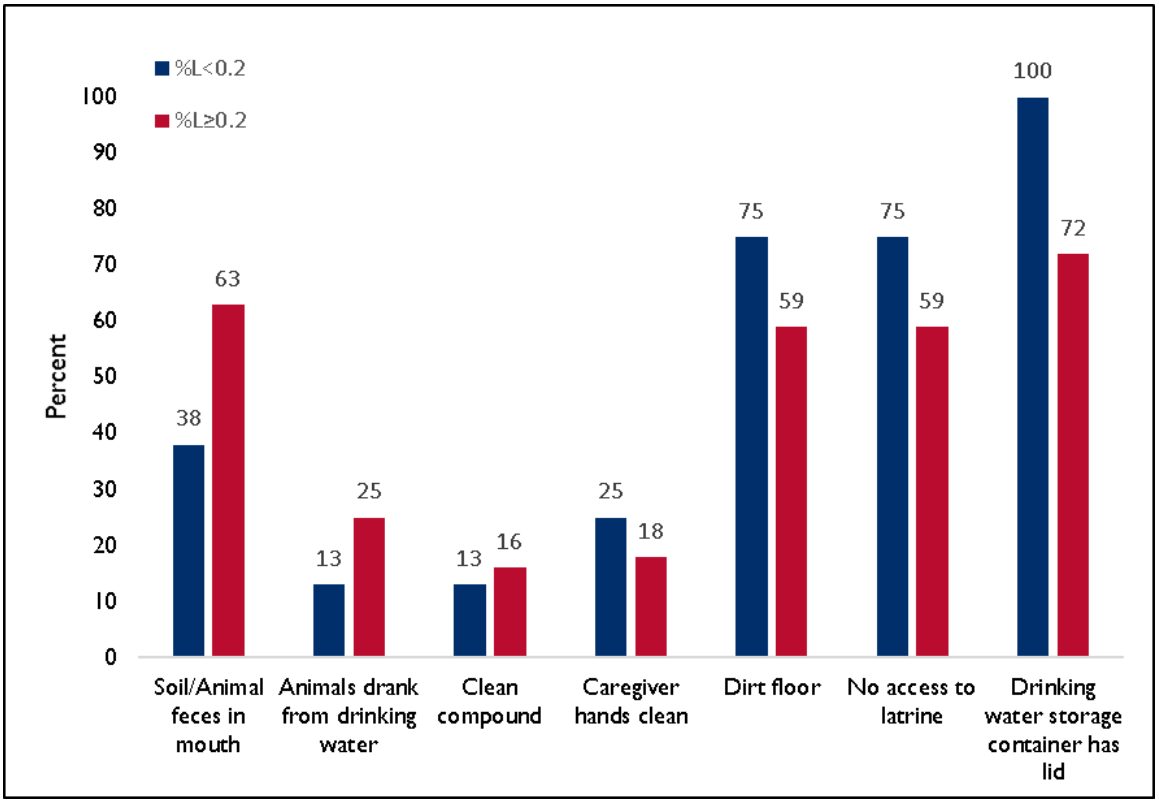


Note: Day 2 is at enrollment into the supplementary feeding program.

Hypothesis: Household WASH practices related to MAM children are associated with EED.

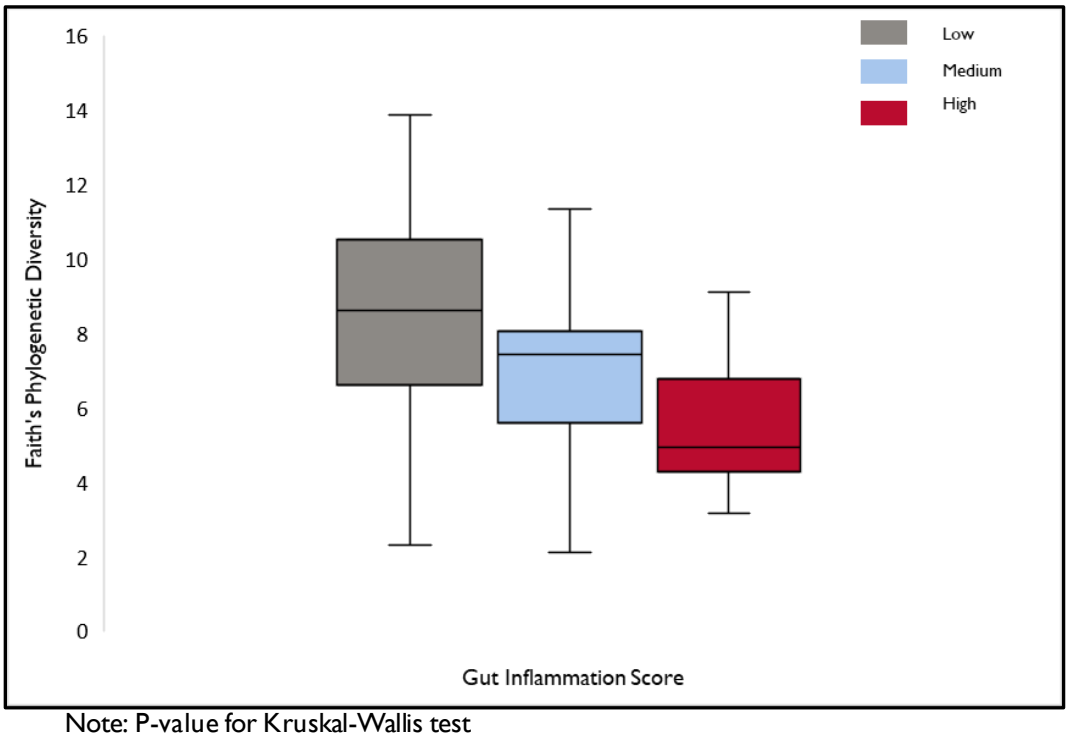
A few differences in WASH practices among study participants with (%L≥0.2, n=32) and without (%L<0.2, n=8) EED were observed **(Figure 8)**. Thirty eight percent of participants without EED were observed putting soil or animal feces in their mouth compared to 63 percent of children with EED. Animals were observed drinking from household drinking water in 13 percent of homes without EED compared to 25 percent of homes with EED. Drinking water storage containers had a lid in homes of all participants without EED compared to 72 percent of homes of participants with EED.

Figure 8. WASH Practices of Participants by %L Status



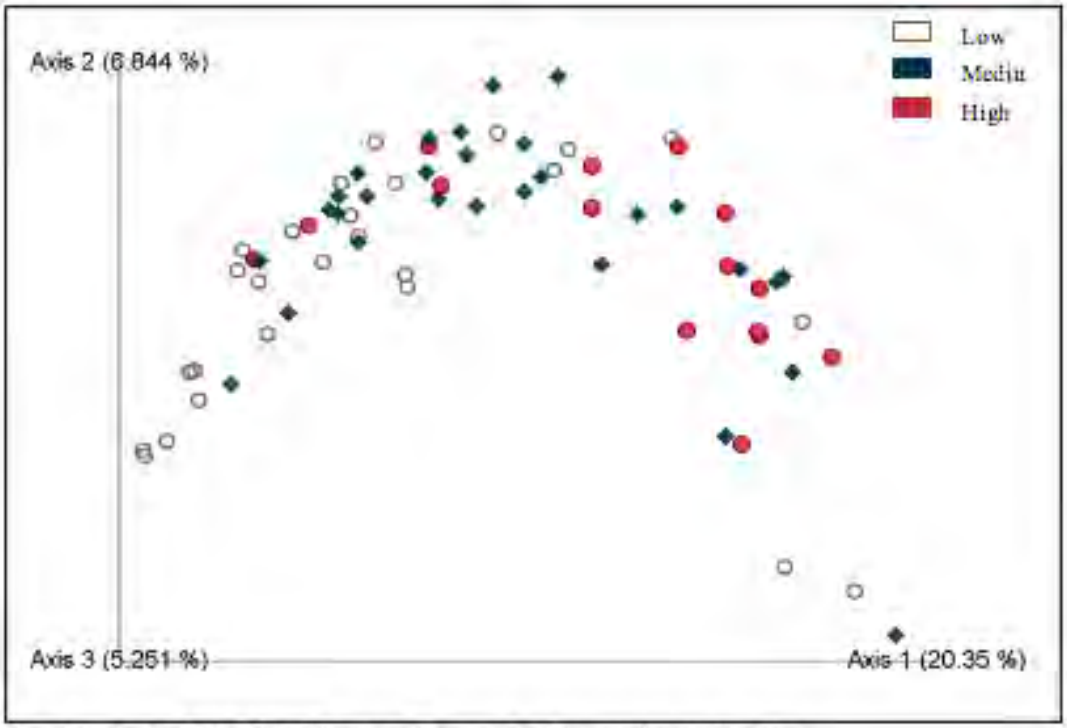
Hypothesis: The microbiota profile of MAM children is associated with EED. There was a statistically significant difference in alpha diversity using Faith's phylogenetic diversity between children of different GIS terciles (p=0.005), borderline for GDS terciles (p=0.052), but not for %L terciles, GDS terciles, or EPS terciles **(Figure 9)**. There was also a statistically significant difference in beta diversity by unweighted UniFrac distance between children of different GIS terciles (p=0.009) **(Figure 10)** and GDS terciles (p=0.048), but not for %L terciles, GSS terciles, or EPS terciles. Statistical significance in alpha diversity and beta diversity between MAM children and their healthy peers was borderline (p=0.059 and p=0.051 respectively).

Figure 9. Alpha Diversity as Measured by Faith's Phylogenetic Diversity for GIS



Note: P-value for Kruskal-Wallis test

Figure 10. Principal Coordinates Analysis Plots for GIS Terciles



Note: P-value for PERMANOVA testing if centroids of the terciles are the same

5. Challenges and Limitations

Challenges: The team faced a number of challenges while implementing the EED sub-study in Pujehun district, which is a resource-constrained setting. Maintaining temperature control was critical for the stool samples. Due to lack of electricity at the study site, stool samples were stored in liquid nitrogen until they could be transferred to a -80°C freezer for long-term storage. When the only liquid nitrogen machine in Sierra Leone broke down, dry ice had to be imported. Importing dry ice not only was expensive, but also made the study reliant on the often unpredictable international flight schedule. However, none of the samples was compromised due to flight delays.

Given the mobile nature of the *Four Foods MAM Treatment Study* and subsequently the EED sub-study, the field team spent only one day at each PHU. Therefore, the sub-study was only able to attempt sample collection one time for each participant. Collecting both stool and urine on the same day was also challenging, because attaching the diaper and urine bag at the same time obstructed the urine bag on a number of occasions.

Limitations: Conducting the L:M test on the same day as stool collection might have affected the AAT protein concentrations. This is because lactulose is a laxative that could have resulted in participants passing watery stool. Stool samples with high water content are expected to show low AAT values (Crossley & Elliott, 1977). Despite the rather low AAT concentrations observed in this sub-study, the values were higher than the mean AAT concentration of 597 ng/mL reported by the Malnutrition and Enteric Disease (MAL-ED) Peru birth cohort (n=303) (Kosek et al., 2013).

Although it was not possible to extend the duration of the second collection of stool samples, a longer exposure period beyond four weeks might have allowed us to see the impact of the study foods on intestinal health. It was not possible to extend the second sample collection to, for example, recovery because this outcome varied from 2 to 12 weeks among the population and we would not be able to control for endogenous differences between children who recovered faster and those who recovered more slowly. Furthermore, conducting all three biomarker assessments at both time points could have allowed us to examine a more complete picture of EED after the intervention. However, this was not possible due to logistical and budgetary constraints.

6. Summary of Findings

The prevalence of EED assessed using %L, MPO, and NEO was high among children with MAM enrolled in a supplemental feeding program in the Pujehun district of Sierra Leone. Three EED scores based on 15 fecal host mRNA transcripts were used: GIS, GSS, and GDS. The EPS, GIS, GSS, and GDS did not correlate with %L. That said, a weak but significant correlation was found between EPS and GIS, suggesting that they are both markers of inflammation, which is characteristic of EED. Previous studies have found mixed results for the association between proteins comprising EPS and L:M ratio or %L (Harper, Mutasa, Prendergast, Humphrey, & Manges, 2018). However, eight fecal host mRNA transcripts were able to identify the presence of EED and severe EED using %L with high sensitivity and specificity. This finding is in contrast to what we found with the factor analysis-based scores (GIS, GSS, and GDS), probably because these scores measure specific domains of EED separately while the eight mRNA transcripts together capture many domains of EED. These eight transcripts were not the same as those reported by a previous study that successfully identified children with severe EED (L:M ratio ≥0.45), also using the random forest classification method (Ordiz et al., 2016).

The level of EED at enrollment that was assessed using any of the four biomarkers of EED did not influence the effectiveness of the study foods except for GDS. More children with high GDS (a sign of good gut health) at enrollment recovered compared to children with lower GDS. This finding suggests that children who start the program with a healthier small intestine are more likely to recover than children with a less healthy small intestine. We might not have found a difference in the effects of the foods at different levels of EED because the foods, though variable in composition, performed comparably in terms of their effect on recovery from MAM. The fact that more children with a healthier small intestine recover was also supported by EPS, where more participants with low scores (sign of less inflammation) graduated from the treatment program compared to participants with a higher score. The difference was not statistically significant (p=0.125), possibly due to the sample size (n=189). Similar to our findings, a study from Bangladesh also reported that effectiveness of a zinc supplementation trial did not vary by presence of EED based on L:M ratio at enrollment (Long et al., 2019).

We did not find an effect of the study foods on EED using 15 fecal host mRNA transcripts. This might be because the study foods do not improve EED or because the exposure period of four weeks was not sufficient to result in significant change. However, we did see trends in certain mRNA transcripts that we would expect to change after an intervention. For example, the concentration of mRNA transcript S100A8[11] was, on average, lower after four weeks on the study foods, signaling reduction in inflammation. Studies that have shown improvements in %L after a nutrition intervention have had a longer exposure period, such as three months or at least eight weeks (Agapova et al., 2018; Cheng et al., 2019).

Household WASH practices were generally poor among participants and caregivers of the EED sub-study. Differences in practices between participants with and without EED (based on %L)

[11] S100A8 encodes the inflammatory protein calprotectin [41].

suggest that interventions targeted at improving these behaviors might be necessary to improve EED in contexts of high risk of MAM. A study from Uganda showed that unsafe drinking water was associated with EED and poor child growth (Lauer et al., 2018). Differences in the microbiota profile of MAM children with varying levels of EED as assessed using different EED biomarkers was statistically significant for GIS terciles when examining both alpha and beta diversity. A study conducted among Malawian children without acute malnutrition found significant differences in beta diversity but not alpha diversity when assessing different degrees of EED based on the L:M ratio (Ordiz et al., 2017).

7. Recommendations

Based on results from the EED sub-study, a few recommendations can be made for future food assistance programs:

- MAM children who start the program with a healthier small intestine (based on GDS) are more likely to graduate from the treatment program within 12 weeks. Strategies to improve intestinal health through efforts in addition to SNFs (such as WASH interventions) should therefore be explored. None of the study foods performed better than CSB+ with oil in enabling MAM children with poor intestinal health to graduate from the program. This sub-study does not support choosing different foods for treatment of MAM children based on EED status at enrollment.
- WASH practices among MAM children and their caregivers were poor. Some WASH practices may be associated with EED, a marker of poor intestinal health. To address EED, it might improve effectiveness to incorporate WASH actions in MAM treatment programs and counsel caregivers on the benefits of good WASH practices.

Acknowledgements
This study would not be possible without the support of the USAID Bureau for Humanitarian Assistance (BHA) and the legacy Office of Food for Peace (FFP) and their ongoing commitment to improving Title II programming in order to address food insecurity in vulnerable populations. The time and support offered to the Food Aid Quality Review by Title II awardees, both headquarters and field staff, as well as Food for Peace Officers abroad and in Washington, D.C. and the Policy and Technical Division of FFP headquarters were also invaluable to informing this study.

References

Agapova, S., Stephenson, K., Manary, M., Weisz, A., Tarr, P. I., Mkakosya, R., Shaikh, N. (2013). Detection of low-concentration host mRNA transcripts in Malawian children at risk for environmental enteropathy. *J Pediatr Gastroenterol Nutr, 56*(1), 66-71. doi:10.1097/MPG.0b013e31826a107a

Agapova, S. E., Stephenson, K. B., Divala, O., Kaimila, Y., Maleta, K. M., Thakwalakwa, C., Manary, M. J. (2018). Additional Common Bean in the Diet of Malawian Children Does Not Affect Linear Growth, but Reduces Intestinal Permeability. *J Nutr, 148*(2), 267-274. doi:10.1093/jn/nxx013

Arndt, M. B., Richardson, B. A., Ahmed, T., Mahfuz, M., Haque, R., John-Stewart, G. C., Walson, J. L. (2016). Fecal Markers of Environmental Enteropathy and Subsequent Growth in Bangladeshi Children. *Am J Trop Med Hyg, 95*(3), 694-701. doi:10.4269/ajtmh.16-0098

Beckmann G, R. A. (2000). *Microbiology of the Intestine*. Hannover: Schluttersche.

Campbell, R. K., Schulze, K. J., Shaikh, S., Mehra, S., Ali, H., Wu, L., Christian, P. (2017). Biomarkers of Environmental Enteric Dysfunction Among Children in Rural Bangladesh. *J Pediatr Gastroenterol Nutr, 65*(1), 40-46. doi:10.1097/mpg.0000000000001557

Campbell, R. K., Schulze, K. J., Shaikh, S., Raqib, R., Wu, L. S. F., Ali, H., Christian, P. (2018). Environmental enteric dysfunction and systemic inflammation predict reduced weight but not length gain in rural Bangladeshi children. *Br J Nutr, 119*(4), 407-414. doi:10.1017/s0007114517003683

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A, 108 Suppl 1*, 4516-4522. doi:10.1073/pnas.1000080107

Cheng, W. D., Wold, K. J., Bollinger, L. B., Ordiz, M. I., Shulman, R. J., Maleta, K. M., Trehan, I. (2019). Supplementation With Lactoferrin and Lysozyme Ameliorates Environmental Enteric Dysfunction: A Double-Blind, Randomized, Placebo-Controlled Trial. *Am J Gastroenterol, 114*(4), 671-678. doi:10.14309/ajg.0000000000000170

Crane, R. J., Jones, K. D., & Berkley, J. A. (2015). Environmental enteric dysfunction: an overview. *Food Nutr Bull, 36*(1 Suppl), S76-87. doi:10.1177/15648265150361s113

Crossley, J. R., & Elliott, R. B. (1977). Simple method for diagnosing protein-losing enteropathies. *Br Med J, 1*(6058), 428-429.

Denno, D. M., VanBuskirk, K., Nelson, Z. C., Musser, C. A., Hay Burgess, D. C., & Tarr, P. I. (2014). Use of the lactulose to mannitol ratio to evaluate childhood environmental enteric dysfunction: a systematic review. *Clin Infect Dis, 59 Suppl 4*, S213-219. doi:10.1093/cid/ciu541

George, C. M., Oldja, L., Biswas, S., Perin, J., Lee, G. O., Kosek, M., Faruque, A. G. (2015). Geophagy is associated with environmental enteropathy and stunting in children in rural Bangladesh. *Am J Trop Med Hyg, 92*(6), 1117-1124. doi:10.4269/ajtmh.14-0672

George, C. M., Oldja, L., Biswas, S. K., Perin, J., Lee, G. O., Ahmed, S., Faruque, A. G. (2015). Fecal Markers of Environmental Enteropathy are Associated with Animal Exposure and Caregiver Hygiene in Bangladesh. *Am J Trop Med Hyg, 93*(2), 269-275. doi:10.4269/ajtmh.14-0694

Harper, K. M., Mutasa, M., Prendergast, A. J., Humphrey, J., & Manges, A. R. (2018). Environmental enteric dysfunction pathways and child stunting: A systematic review. *PLoS Negl Trop Dis, 12*(1), e0006205. doi:10.1371/journal.pntd.0006205

Hossain, M. I., Nahar, B., Hamadani, J. D., Ahmed, T., Roy, A. K., & Brown, K. H. (2010). Intestinal mucosal permeability of severely underweight and nonmalnourished Bangladeshi children and effects of nutritional rehabilitation. *J Pediatr Gastroenterol Nutr, 51*(5), 638-644. doi:10.1097/MPG.0b013e3181eb3128

Keusch, G. T., Denno, D. M., Black, R. E., Duggan, C., Guerrant, R. L., Lavery, J. V., Brewer, T. (2014). Environmental enteric dysfunction: pathogenesis, diagnosis, and clinical consequences. *Clin Infect Dis, 59 Suppl 4*, S207-212. doi:10.1093/cid/ciu485

Kosek, M., Haque, R., Lima, A., Babji, S., Shrestha, S., Qureshi, S., Gottlieb, M. (2013). Fecal markers of intestinal inflammation and permeability associated with the subsequent acquisition of linear growth deficits in infants. *Am J Trop Med Hyg, 88*(2), 390-396. doi:10.4269/ajtmh.2012.12-0549

Kosek, M. N. (2017). Causal Pathways from Enteropathogens to Environmental Enteropathy: Findings from the MAL-ED Birth Cohort Study. *EBioMedicine, 18*, 109-117. doi:10.1016/j.ebiom.2017.02.024

Lauer, J. M., Duggan, C. P., Ausman, L. M., Griffiths, J. K., Webb, P., Bashaasha, B., Ghosh, S. (2018). Unsafe Drinking Water Is Associated with Environmental Enteric Dysfunction and Poor Growth Outcomes in Young Children in Rural Southwestern Uganda. *Am J Trop Med Hyg, 99*(6), 1606-1612. doi:10.4269/ajtmh.18-0143

Ledjeff, E., Artner-Dworzak, E., Witasek, A., Fuchs, D., & Hausen, A. (2001). Neopterin Concentrations in Colon Dialysate *Pteridines* (Vol. 12, pp. 155).

Lin, A., Ali, S., Arnold, B. F., Rahman, M. Z., Alauddin, M., Grembi, J., Luby, S. P. (2019). Effects of water, sanitation, handwashing, and nutritional interventions on environmental enteric dysfunction in young children: a cluster-randomized controlled trial in rural Bangladesh. *Clin Infect Dis*. doi:10.1093/cid/ciz291

Lin, A., Arnold, B. F., Afreen, S., Goto, R., Huda, T. M., Haque, R., Luby, S. P. (2013). Household environmental conditions are associated with enteropathy and impaired growth in rural Bangladesh. *Am J Trop Med Hyg, 89*(1), 130-137. doi:10.4269/ajtmh.12-0629

Long, J. M., Mondal, P., Westcott, J. E., Miller, L. V., Islam, M. M., Ahmed, M., Krebs, N. F. (2019). Zinc Absorption from Micronutrient Powders Is Low in Bangladeshi Toddlers at Risk of Environmental Enteric Dysfunction and May Increase Dietary Zinc Requirements. *J Nutr, 149*(1), 98-105. doi:10.1093/jn/nxy245

Lozupone, C. A., & Knight, R. (2008). Species divergence and the measurement of microbial diversity. *FEMS Microbiol Rev, 32*(4), 557-578. doi:10.1111/j.1574-6976.2008.00111.x

Mahfuz, M., Das, S., Mazumder, R. N., Masudur Rahman, M., Haque, R., Bhuiyan, M. M. R., Ahmed, T. (2017). Bangladesh Environmental Enteric Dysfunction (BEED) study: protocol for a community-based intervention study to validate non-invasive biomarkers of environmental enteric dysfunction. *BMJ Open, 7*(8), e017768. doi:10.1136/bmjopen-2017-017768

Morita, T., Perin, J., Oldja, L., Biswas, S., Sack, R. B., Ahmed, S.,George, C. M. (2017). Mouthing of Soil Contaminated Objects is Associated with Environmental Enteropathy in Young Children. *Trop Med Int Health, 22*(6), 670-678. doi:10.1111/tmi.12869

Ordiz, M. I., Davitt, C., Stephenson, K., Agapova, S., Divala, O., Shaikh, N., & Manary, M. J. (2018). EB 2017 Article: Interpretation of the lactulose:mannitol test in rural Malawian children at risk for perturbations in intestinal permeability. *Exp Biol Med (Maywood), 243*(8), 677-683. doi:10.1177/1535370218768508

Ordiz, M. I., Shaikh, N., Trehan, I., Maleta, K., Stauber, J., Shulman, R.,Manary, M. J. (2016). Environmental Enteric Dysfunction Is Associated With Poor Linear Growth and Can Be Identified by Host Fecal mRNAs. *J Pediatr Gastroenterol Nutr, 63*(5), 453-459. doi:10.1097/mpg.0000000000001315

Ordiz, M. I., Stephenson, K., Agapova, S., Wylie, K. M., Maleta, K., Martin, J., Manary, M. J. (2017). Environmental Enteric Dysfunction and the Fecal Microbiota in Malawian Children. *Am J Trop Med Hyg, 96*(2), 473-476. doi:10.4269/ajtmh.16-0617

Ordiz, M. I., Wold, K., Kaimila, Y., Divala, O., Gilstrap, M., Lu, H. Z., & Manary, M. J. (2018). Detection and interpretation of fecal host mRNA in rural Malawian infants aged 6-12 months at risk for environmental enteric dysfunction. *Exp Biol Med (Maywood), 243*(12), 985-989. doi:10.1177/1535370218794418

Prendergast, A. J., Humphrey, J. H., Mutasa, K., Majo, F. D., Rukobo, S., Govha, M., Stoltzfus, R. J. (2015). Assessment of Environmental Enteric Dysfunction in the SHINE Trial: Methods and Challenges. *Clin Infect Dis, 61 Suppl 7*, S726-732. doi:10.1093/cid/civ848

Saiki, T. (1998). Myeloperoxidase concentrations in the stool as a new parameter of inflammatory bowel disease. *Kurume Med J, 45*(1), 69-73.

Shulman, R. J., Schanler, R. J., Lau, C., Heitkemper, M., Ou, C. N., & Smith, E. O. (1998). Early feeding, antenatal glucocorticoids, and human milk decrease intestinal permeability in preterm infants. *Pediatr Res, 44*(4), 519-523. doi:10.1203/00006450-199810000-00009

Stauber, J., Shaikh, N., Ordiz, M. I., Tarr, P. I., & Manary, M. J. (2016). Droplet digital PCR quantifies host inflammatory transcripts in feces reliably and reproducibly. *Cell Immunol, 303*, 43-49. doi:10.1016/j.cellimm.2016.03.007

Subramanian, S., Huq, S., Yatsunenko, T., Haque, R., Mahfuz, M., Alam, M. A., Gordon, J. I. (2014). Persistent gut microbiota immaturity in malnourished Bangladeshi children. *Nature, 510*(7505), 417-421. doi:10.1038/nature13421

Yu, J., Ordiz, M. I., Stauber, J., Shaikh, N., Trehan, I., Barnell, E., Manary, M. J. (2016). Environmental Enteric Dysfunction Includes a Broad Spectrum of Inflammatory Responses and Epithelial Repair Processes. *Cell Mol Gastroenterol Hepatol, 2*(2), 158-174.e151. doi:10.1016/j.jcmgh.2015.12.002

Appendix 1: Description of Fecal Host mRNA Transcripts

Gene symbol	Description	EED domains	Functions (Source: https://www.genecards.org)
AQP9	Aquaporin 9	Absorption Immune response	Membrane channels that allow passage of noncharged particles; might be involved in immune response
BIRC3	Baculoviral IAP Repeat Containing 3	Immune response	Prevents apoptosis (cell death) by binding to tumor necrosis factor receptor-associated factors
CD53	CD53 Molecule	Immune response Cell adhesion Cell development	Cell surface protein involved in cell growth and development; binds integrins; lack of this protein results in immunodeficiency
CDX1	Caudal Type Homeobox 1	Cell differentiation	Regulates differentiation of intestinal cells
DECR1	2,4-Dienoyl-CoA Reductase 1	Fatty acid metabolism	Enzyme involved in fatty acid metabolism
DEFA6	Defensin, Alpha 6, Paneth Cell-Specific	Immune response	Antimicrobial and cytotoxic peptides present in neutrophils (innate immune system) and mucosal surfaces; might protect cells against HIV-1
HLA-DRA	Major Histocompatibility Complex, Class II, DR Alpha	Immune response	Expressed on the surface of antigen-presenting cells, binds peptides to present to T cells (adaptive immune system)
IFI30	Lysosomal Thiol Reductase	Immune response	Enzyme that has a role in antigen processing (adaptive immune response)
LYZ	Lysozyme	Immune response	Antimicrobial agent associated with monocyte-macrophage system (innate immune system)

Gene symbol	Description	EED domains	Functions (Source: https://www.genecards.org)
MUC12	Mucin 12	Permeability Cell differentiation	Formation of protective mucous barrier and differentiation of epithelial cells
PIK3AP1	Phosphoinositide-3-Kinase Adaptor Protein 1	Immune response	Adaptor involved in B cell development (adaptive immune system), links Toll-like receptors to PIK3 to prevent excessive production of cytokines.
REG1A	Regenerating Islet-Derived 1 Alpha	Repair/Injury	Regeneration of islet cells of the pancreas
REG3A	Regenerating Islet-Derived 3 Alpha	Immune response	Mediates bacterial killing; pancreatic secretary protein possibly involved in cell proliferation and/or differentiation
S100A8	S100 Calcium Binding Protein A8 (Calprotectin)	Immune response	Encodes the protein calprotectin, which is a marker of neutrophil activity (innate immune system)
SELL	Selectin L	Immune response Cell adhesion	Cell surface protein involved in binding leucocytes (innate immune response)

USAID
FROM THE AMERICAN PEOPLE

Comparative Cost-Effectiveness of Four Supplementary Foods in Treating Moderate Acute Malnutrition in Children 6-59 Months in Sierra Leone

Section 4: Neurocognitive Sub-Study

A Report from the Food Aid Quality Review

Prepared by:

Jukka Lappanen
Mark J Manary
Claire Godbout
Stacy Griswold
Breanne Langlois
Ye Shen
Ilana Cliffer
Donna Wegner
Devika Suri
Kenneth Chui
Irwin Rosenberg
Patrick Webb
Beatrice Lorge Rogers

JANUARY 2021

This report was made possible by the generous support of the American people through the support of the United States Agency for International Development's Bureau for Humanitarian Assistance (USAID/BHA) and the legacy Office of Food for Peace (FFP) under the terms of Contract AID-OAA-C-16-00020, managed by Tufts University.

The contents are the responsibility of Tufts University and its partners in the Food Aid Quality Review (FAQR) and do not necessarily reflect the views of USAID or the United States Government.

The authors have no conflict of interest to declare.

Recommended Citation

Lappanen, Jukka; Manary, Mark J; Godbout, Claire; Griswold, Stacy; Langlois, Breanne; Shen, Ye; Cliffer, Ilana; Wegner, Donna; Suri, Devika; Chui, Kenneth; Rosenberg, Irwin; Webb, Patrick; Rogers, Beatrice. 2020. *Comparative Cost-effectiveness of Four Supplementary Foods in Treating Moderate Acute Malnutrition in Children 6-59 Months in Sierra Leone- Section 4: Neurocognition Substudy*, Report to USAID from the Food Aid Quality Review. Boston, MA: Tufts University.

For correspondence, contact:

Beatrice Lorge Rogers, PhD
Friedman School of Nutrition Science and Policy
Tufts University
150 Harrison Avenue
Boston, MA 02111
beatrice.rogers@tufts.edu
http://www.foodaidquality.org

ABBREVIATIONS AND ACRONYMS

CSB+ w/oil	Corn-Soy Blend Plus and Fortified Vegetable Oil
CSWB w/oil	Corn-Soy Whey Blend and Fortified Vegetable Oil
MAM	Moderate Acute Malnutrition
MUAC	Mid-Upper Arm Circumference
PHU	Peripheral Health Units
RUSF	Ready-to-Use Supplementary Food
SC+A	Super Cereal Plus Amylase
SRT	Saccadic Reaction Time

TABLE OF CONTENTS

Section 4: Neurocognition Sub-study

ABSTRACT

Background: Moderate acute malnutrition (MAM) is a potentially life-threatening condition that affects the development and function of multiple organ systems, including the brain. The assessment of functional deficits in MAM has remained difficult given the lack of objective and widely-applicable tests. This study examined the feasibility, measurement properties, and utility of a novel eye-tracking test in the assessment of neurocognitive function in young children in low-resource settings. The study also tested the hypothesis that MAM is associated with impaired neurocognitive functions that are improved after treatment with supplementary food.

Methods: The study used a case-control (4:1) design to compare neurocognitive function in children with MAM and well-nourished controls. The data were collected as a part of a larger intervention study comparing the cost-effectiveness of four supplementary foods in the treatment of MAM in Pujehun District, Sierra Leone. Infants aged 7- to 11-months diagnosed with MAM (N=131) and same-age controls without MAM (N = 60) were enrolled in the study. They were tested with eye tracking and conventional observational tests of visual orienting. The tests were administered upon enrollment to the study/feeding program (i.e., Pretest/Visit 1) and again four weeks later (Post-test/Visit 2). The primary outcome was saccadic reaction time (SRT), reflecting the speed of visual orienting to stimuli on a computer screen. Secondary outcomes were measures reflecting the spatial and temporal characteristics of visual fixations while the infant was exploring a dynamic social scene.

Results: Data on visual orienting were obtained for 76 percent of the infants at the first attempt of eye tracking. Visual orienting responses to stimuli on a computer screen were frequent in MAM and control children (Mean = 95%), yielding sufficient data for reliable estimation of SRTs (test-retest r = .60, P<0.001) and the mean duration of visual fixations (test-retest r = .53, P<0.001). MAM children had marginally slower SRTs than controls in the baseline assessment. SRT improved among the children with MAM after four weeks of supplementary feeding (-6 ms $\pm$ 2, P=0.006), as did fixation duration (+34 ms $\pm$ 6, P<.0001). For control children, there were no significant changes in SRT (+4ms $\pm$ 3, P=0.19) or fixation duration (+8ms $\pm$9, P=0.43). No group or treatment effects were found in the conventional observational tests of visual orienting.

Conclusions: The results demonstrate the promise of eye tracking in the assessment of neurocognitive status in young, vulnerable groups in field settings and show unique evidence for the functional benefits of supplementary feeding of children with MAM.

2. Background

Human brain development is dependent on access to a supportive environment with adequate levels of nutrition, cognitive stimulation, and responsive caregiving (Georgieff, 2007; Nelson et al., 2006). Children who lack access to this optimum due to malnutrition and other forms of adversity are at heightened risk for compromised development of elementary neurocognitive capacities in early childhood, and persistent emotional and social problems in adulthood (Lu et al., 2016; Walker et al., 2011). A timely intervention that supplies the developing brain with adequate nutrition may be effective in preventing such long-term negative consequences (Prado & Dewey, 2014), but studies on the neurocognitive benefits of supplementary feeding programs are scarce given the lack of standardized and widely-applicable measures. It is increasingly recognized, however, that an important next step in the evaluation of the treatment protocols for early childhood malnutrition is to extend the measurement of treatment outcomes from anthropometric indicators to other measures of the child's functional status, including neurocognitive function (Hsieh et al., 2016; Suri & Rosenberg, 2018).

Neurocognitive Development and Nutritional Deficiencies

The human brain undergoes rapid growth and development in the last trimester of pregnancy and the early postnatal period (Georgieff, 2007). This development results in the appearance of a repertoire of elementary neurocognitive capacities that: i) help the child to orient attention to important sources of information in the physical and social environment (e.g., parents' faces); ii) facilitate early learning (e.g., detection of statistical regularities in the environment); and iii) serve as "scaffolds" in the acquisition of more complex cognitive and social skills (Rose et al., 2003).

The development of elementary neurocognitive capacities in infants is based on activity-dependent refinement of connections in distributed neural networks. For example, the appearance of stereopsis (i.e., use of visual input from the two eyes in depth perception, Jando et al., 2012) is based on the formation of distinct columns of neurons in the occipital neocortex that respond selectively to input from one eye (Espinosa & Stryker, 2012). Similarly, age-related changes in attention to faces in infants (Leppänen, 2017) are likely to be related to the appearance of distinct populations of face-sensitive neural cells in the occipital and temporal areas of the brain (Arcaro et al., 2017). In addition to the appearance of new abilities, early neurocognitive development in infants is characterized by rapid age-related improvements in the efficiency of neurocognitive functions, such as the speed of visual orienting (Alahyane et al., 2016). These changes are likely to be related to myelination (i.e., formation of lipid-rich sheaths around neuronal axons) and associated improvements in the neural conduction velocity. In the visual system, myelination first appears in optic radiations (starting around three to four months after birth) and subsequently in occipital and parietal cortices (starting around four to six months after birth [Deoni et al., 2012]).

Nearly all aspects of the early neurodevelopment in infants are vulnerable to nutritional deficiencies (Georgieff, 2007; Prado & Dewey, 2014). While the mechanisms of this vulnerability are poorly understood, the neurodevelopmental effects of malnutrition are hypothesized to be based on: i) specific impairments arising from a direct dependency of

a neural or developmental process on a particular nutrient (e.g., several studies have examined links between visual brain systems and long-chain polyunsaturated fatty acids [Birch et al., 2010; Brenna, 2011; Colombo et al., 2013; 2016; Diau et al., 2005; El-Khayat et al., 2007; Gould et al., 2014; Westerberg et al., 2011]); or ii) a more diffuse effect arising from many neurodevelopmental processes being sensitive to nutritional deficiencies, including protein-energy malnutrition, as well as deficiencies in fatty acids, iron, iodine and thyroid hormones, zinc, choline, and vitamin B (for reviews, see Georgieff, 2007; Prado & Dewey, 2014). The latter may explain the findings of widespread neurocognitive and behavioral problems in individuals with a history of acute malnutrition in early childhood, including increased susceptibility to distraction, frustration, and achievement problems at school age, and intellectual and emotional dysfunction in adulthood (Galler et al, 1983; 1989; Peter et al., 2016).

Given that early neurocognitive-development is highly dependent on activity and relies on the child's ability to interact with the physical and social environment, neurodevelopmental effects in malnourished children may also arise from the fact that this condition is often characterized by reduced interest in the environment, reduced motor activity and reduced exploratory behavior (Prado & Dewey, 2014). If a child is less active, inactivity may inadvertently lead to a negative cycle in which the caregiver becomes less engaged with the child in joint activities that stimulate the child cognitively and offer opportunities for learning (Prado & Dewey, 2014).

Assessment of Neurocognitive Development

The neurodevelopmental effects of malnutrition may be mitigated by timely interventions that supply the developing brain with required nutrients. However, there is a scarcity of studies on the neurocognitive benefits of supplementary feeding programs in food-insecure environments (Prado & Dewey, 2014), mostly reflecting the difficulties in measuring children's neurocognitive function under field conditions. Traditional observational tests of infant neurocognition, for example, are impractical for repeated use in routine clinical or field settings, rely heavily on user experience and education, and are insensitive to child developmental trajectories before the second year of life (Nishimura et al., 2016).

In recent years, the opportunities to examine neurocognitive function in infants have improved with the advent of technologies that allow for computerized tracking of infant eye movements as well as full automation of procedures that have long been manual, laborious and error-prone (Jones et al., 2014; Leppänen et al., 2015). These technologies have been mostly used in academic research (although see Forssman et al., 2017, Hernik et al., 2018, Pyykkö et al., 2018), but there are growing hopes that some of the new technologies can be developed into practical field-tests for collecting data on specific aspects of human behavior (Itti, 2015; Tseng et al., 2012) and, as such, would also help in creating new, scalable methods for monitoring changes in neurocognitive function in children (Jones et al., 2014; Leppänen et al., 2015).

The technological advancements in the measurement of eye movements are of interest because different metrics of eye movements may offer reliable and widely-applicable

indicators of brain function and its problems across the human lifespan. Eye movement control develops rapidly within the first months of life and gives infants an ability to look at things that are of interest to them, keep their eyes focused on objects and prevent their eyes from moving to things that are distracting (Leppänen, 2017). These abilities are controlled by a distributed network of subcortical and cortical areas (Munoz et al., 2004).

Several studies have further shown that the timing of human eye movements is highly sensitive to changes in a person's neurocognitive status and to individual differences in neurocognitive capacity. Saccadic eye movements (i.e., rapid movements of eyes from one location to another) become slower after a mild traumatic brain injury and subsequently return to the pre-event baseline as the person recovers from the injury (e.g., Pearson et al., 2007). There are also data showing that individual differences in the latency of saccadic eye movements and the mean duration of periods of stationary attentional focus (fixations) in infancy are associated with scores on the conventional measures of cognitive and behavioral capacities in mid childhood (Dougherty & Haith, 1997; Papageorgiou et al., 2014; Rose et al, 2012).

While infant eye movements are easy to measure and correlate with neurocognitive status, a number of questions remain regarding the use of this measure in the assessment of infants and children in low-resource settings. First, the technical feasibility of repeated eye-tracking tests has not been systematically examined in low-resource settings and in children with MAM. Second, the measurement properties of repeated eye-tracking tests are still poorly understood (i.e., variability of test scores across different tests), although existing evidence from high-resource laboratories points to relatively high stability of test results (Leppänen et al., 2015). Limited variability of test scores in populations that are not undergoing changes in neurocognitive status (e.g., well-nourished children without neurological abnormalities) would support the ability of the test to detect changes in neurocognitive status. Further research addressing these questions will be important to determine the utility of eye tracking as an outcome in supplemental feeding programs.

3. Aims and Objectives

The Neurocognitive Sub-study had the following aims and hypotheses:

Aim 1: To conduct a pilot study to assess the feasibility, measurement properties, and utility of eye-tracking tests in monitoring neurocognitive changes in 7 and 18-month-old children over a four-week period (i.e., the typical recovery period from MAM in supplementary feeding programs).

Aim 2: To examine the effects of MAM and of treatment with four supplementary foods on neurocognitive function in children between the ages of 7 to 11-months.

Hypothesis 2a: Compared to well-nourished control children, children with MAM will have slower saccadic reaction time and higher incidence of missing saccades (i.e., reduced speed and accuracy of visual orienting) and shorter periods of fixation/attention

hold on visual stimuli (i.e., reduced oculomotor/attention control) in brief eye-tracking tests.

Hypothesis 2b: Treatment with supplementary food will lead to a recovery of neurocognitive function in MAM. Compared to the assessment at enrollment, children with MAM will exhibit shorter saccadic reaction times and longer fixation durations after four weeks of supplementary feeding.

Hypothesis 2c: There are differences in the effects of the four foods on measures of neurocognitive function in 7 to 11-month-old children with MAM after four weeks of supplementation.

Primary Outcomes: The outcomes measured were: i) child nutritional status (measures of mid-upper arm circumference [MUAC]); ii) neurocognitive function, including counts of saccades to high-contrast patterns (consistency and accuracy of visual orienting), mean saccadic reaction time (speed of visual orienting in different conditions), and mean fixation duration.

4. Methods

4.1 Study Site

Data were collected as a part of a larger intervention study comparing the cost-effectiveness of four supplementary foods in the treatment of MAM in Pujehun District, Sierra Leone. The study was implemented in 29 Peripheral Health Units (PHUs), which were randomized to administer one of four supplementary foods to children 6-59 months of age diagnosed with MAM—defined as MUAC ≥ 11.5 cm and < 12.5 cm without bipedal oedema. Caregivers of children admitted into the feeding program were seen every two weeks until an outcome was reached up to seven rations total and received one of the following foods:

1) Super Cereal Plus with amylase (SC+A): 136 g SC+A/day; 2) Corn-Soy Blend Plus and fortified vegetable oil (CSB+ w/oil): 86 g CSB+/day and 26 g fortified vegetable oil/day; 3) Corn-Soy Whey Blend and fortified vegetable oil (CSWB w/oil): 86 g CSWB/day and 26 g fortified vegetable oil/day; and 4) Ready-to-Use-Supplementary Food (RUSF): 100 g RUSF/day. Depending on recovery (MUAC ≥ 12.5cm), children received the ration for two, four, six, eight, ten or twelve weeks. Children who had not recovered after twelve weeks were referred to the clinic or another health facility for further treatment.

The study was piloted in the study area by assessing children with MAM and well-nourished controls from two age groups (7 months ±14 days and 18 months ±14 days old, targeted sample size N = 32/group). The pilot study showed that the eye-tracking tests were feasible in both age groups, although it was also found that a high proportion of children receiving treatment for MAM in the study area were under 12 months. Given these findings and the hypotheses regarding the effects of malnutrition on early brain development, the main neurocognitive sub-study focused on younger infants and a more narrowly specified developmental window (7- to 11-month-old children). Eligibility

criteria were age between 7 and 11 months and presence of MAM, while exclusion criteria were: a) a known history of visual problems; b) any neurological problem[1]; c) congenital malformation; d) participation in feeding program during the past month; and e) ongoing participation in a supplementary feeding program other than the one in the current study.

Children enrolled in the supplementary feeding program and meeting the age-eligibility were invited for further screening for participation in the sub-study. After a child with MAM was recruited, a control who met the age-eligibility criterion (i.e., 7 to 11 months), had MUAC ≥ 12.5 cm, and resided in the area of the same PHU was screened and recruited for the study. The same exclusion as those used with the MAM group were applied for the control group.

Children were tested at enrollment into the supplementary feeding program and again four weeks later. The four-week time period was chosen because all MAM children received four weeks of rations, even if they achieved recovery after two weeks.

The study received ethical approval from the Tufts University Health Sciences Institutional Review Board, the Washington University in St. Louis School of Medicine Institutional Review Board and the Sierra Leone Ethics and Scientific Review Committee. Beneficiary caregivers signed or thumb-printed the consent form on behalf of their infants.

4.2 Sample Size

Of the 202 infants enrolled in the sub-study, 191 participated in the eye-tracking assessments at Visit 1 (baseline) and Visit 2 (four weeks post-enrollment). Of these, three infants (2 percent) were excluded after prescreening due to non-eligibility (neurological history), and 44 infants (23 percent) for insufficient number of valid eye tracking test trials from Visits 1 or 2 (<10 trials/visit, **Table 1**). Demographic and anthropometric data on the sample children (cases and controls) are provided in **Table 1**.

[1] Known neurological problems, visual problems or congenital malformation were screened on the basis of caregiver report.

Table 1. Sample Demographic and Anthropometric Data by Control and Study Arm.

N	Control	CSB+	CSWB	RUSF	SC+A
Enrolled and tested	60	35	20	39	37
Not eligible	0 (0%)	2 (11%)	0 (0%)	1 (5%)	0 (0%)
Insufficient eye-tracking data	14 (23%)	3 (9%)	8 (40%)	8 (21%)	11 (30%)
Final N	46 (77%)	30 (86%)	12 (60%)	30 (77%)	26 (70%)
Mean Age in Days (SD)	274.4 (41.6)	266.6 (44.6)	259.3 (45.2)	259.4 (37.5)	271.5 (42.1)
Gender (Female/Male)	22/24	15/15	8/4	20/10	21/5
Birth weight (kg)	3.2	2.9	3.0	3.0	3.1
Beneficiary caregiver Illiterate	40 (87%)	27 (90%)	12 100%)	24 (80%)	22 (85%)
Visit 1 MUAC (cm)	13.7	11.9	11.9	11.9	12.0
Visit 2 MUAC (cm)	13.9	12.3	12.3	12.2	12.4
Visit 1 Height (cm)	67.6	64.1	64.3	63.8	64.6
Visit 2 Height (cm)	68.3	65.1	65.6	65.0	66.0
Visit 1 Weight (kg)	7.6	6.0	6.1	6.0	6.1
Visit 2 Weight (kg)	7.8	6.5	6.5	6.4	6.6
Visit 1 Head C (cm)	43.5	42.2	42.0	42.1	41.8
Visit 2 Head C (cm)	43.8	42.6	42.9	42.6	42.4

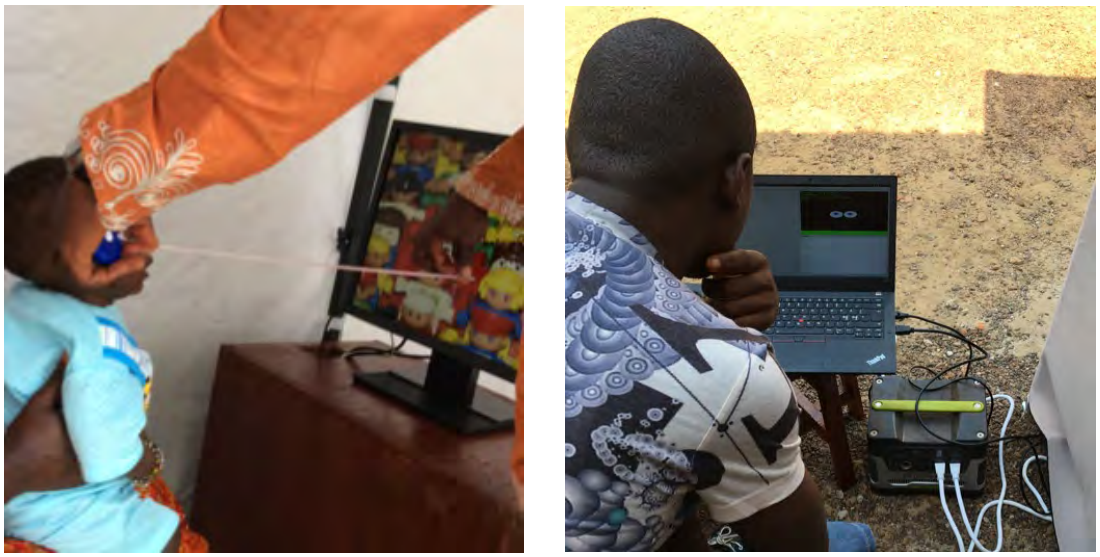
4.3 Data Collection and Management

All participants enrolled in the study were assessed by conventional observational tests and novel eye-tracking-based tests of visual orienting (i.e., saccadic reaction time and fixation measures). The primary outcome was the speed of visual orienting (i.e., saccadic reaction time) as assessed by eye tracking. Secondary outcomes were eye-tracking measures reflecting the spatial and temporal characteristics of visual fixations while the infant was exploring a dynamic social scene.

4.4 Anthropometric Assessment

Infant length was assessed to the nearest 1 mm by using a length board (Harpenden Infantometer, Holtain Limited, Crosswell, Crymych, UK), weights by using an electronic scale (SECA 735, Seca GmbH & Co., Hamburg, Germany) and MUAC and head circumference to the nearest 1 mm with insertion tapes (ShorrTape, Weigh and Measure, LLC, Olney, MD). More detail on the anthropometric measures may be found in the main study section of this report.

Figure 1. Eye Tracking Assessment



A child positioned for eye tracking assessment in a field laboratory tent (the child in the picture is not related to the study). During the assessment, infants viewed short sequences of visual stimuli on the screen (colorful cartoon animations of common objects.) Data collector monitored the assessment outside the laboratory via real-time visualization of eye tracking. The devices used in the testing were powered by a solar-chargeable battery.

4.5 Eye-tracking Testing

Infants were assessed in a 1.5 m (width) x 1.5 m length x 2 m (height) canopy tent with custom-made walls of monochromatic dimout fabric (**Figure 1**). The tent was mounted in the vicinity of a PHU, typically on a covered outdoor porch area. The caregiver held the infant in a baby carrier so that the infant's eyes were facing forward and positioned at approximately 60 cm viewing distance from a 19-Inch monitor (Acer V196L LCD monitor, Resolution: 1280 x 1024, Refresh rate: 60 Hz, Response time: 5 ms) and a screen-based Tobii X3-120 eye tracker (Tobii Technology, Stockholm, Sweden). The monitor and the eye tracker were placed on a wooden desk that was custom-made to be at suitable height for the participants in the sample age range. Infant status during testing was monitored via a video recorder and a baby monitor. The video recorder (Sony HDR-CX240E) was mounted on top of the monitor. The baby monitor (Summer) was placed next to the monitor on the desk.

A Sierra Leonian, Mende-speaking data collector was trained to perform the eye tracking assessments. The data collector instructed the caregiver to look away from the screen and positioned the caregiver, holding the infant in a forward-facing baby carrier, so that the infant's eyes were at an optimal distance from the eye tracker. The data collector then moved outside the tent laboratory to start the tests while constantly monitoring the caregiver and the infant through two monitors. One monitor showed an

overall view of the caregiver and the infant. The other monitor showed a visualization of infant gaze position on the screen.

During the assessment, infants viewed short sequences of visual stimuli on the screen (colorful cartoon animations of common objects). A detailed description of the stimuli is given in **Appendix 1**. While the infants were viewing the stimuli on the screen, their pupil diameter, point of gaze in the display area and position of the eye in a three-dimensional user coordinate system were recorded (120 samples/sec). The testing consisted of two parts, each lasting three to four minutes. If the infant became restless, inattentive or fussy during the procedure or if the eye tracking system lost contact of the eyes, the experimenter paused the test and entered the laboratory to administer a break and/or to perform required adjustments (e.g., adjusted the infant's position).

The presentation of the stimuli on the screen and the synchronization of the data regarding stimulus onset with the eye-tracking data stream was controlled by an open-source software written in Python programming language (https://github.com/infant-cognition-tampere/drop) together with Psychopy functions (Peirce, 2007) and a Tobii SDK plug-in. The software includes a graphical user interface that allowed the data collector to follow real-time visualizations of the stimulus presentation and eye tracking data streams throughout the testing (**Figure 1**). The eye-tracking software was run on Linux operating system (Mint 17) and a Lenovo X260 laptop computer (Lenovo Group Ltd.) and were connected to the stimulus display via a HDMI-DVI cable and to the eye tracker via an Ethernet cable and a TRENDnet 4-Port Broadband Router.

The devices used in the testing were powered by a solar-chargeable battery (GoalZero Yeti 400, NRG Energy, Houston, Texas, USA). A fully-charged battery supplied power for a full day of testing.

4.6 Conventional Eye Movement Testing

Conventional eye testing included tests of saccades, smooth pursuit and fixation (Olsson et al., 2015). In the saccade and smooth pursuit tests, the first target (a colorful rattle) was presented in the child's field of vision at eye level, centered with respect to the child's line of gaze, and at approximately 50 cm from the child's eyes. When the child fixated the target, a second similar target was presented to the child's lateral field, at the same level as the first target and approximately 50 cm to the left or right of the first target. After the child shifted gaze from the first (central) to the second (lateral) target, the lateral target was moved toward the central target. The procedure was repeated twice for both directions. Saccades from the central to the lateral target and smooth pursuit eye movements following the moving target were scored by using a 3-point scale, where 0 indicated absence of saccade or smooth pursuit, 1 partial, and 2 complete saccade or smooth pursuit.

In the tests of visual fixation, the data collector presented a target (toy) for four seconds in an alternating sequence in the child's frontal, lower left, upper left, lower right and

upper right visual field. The child's response (fixation) was scored by using a 3-point scale where 0 indicated no fixation, 1 partial, and 2 complete fixation of the target.

4.7 Data Analysis

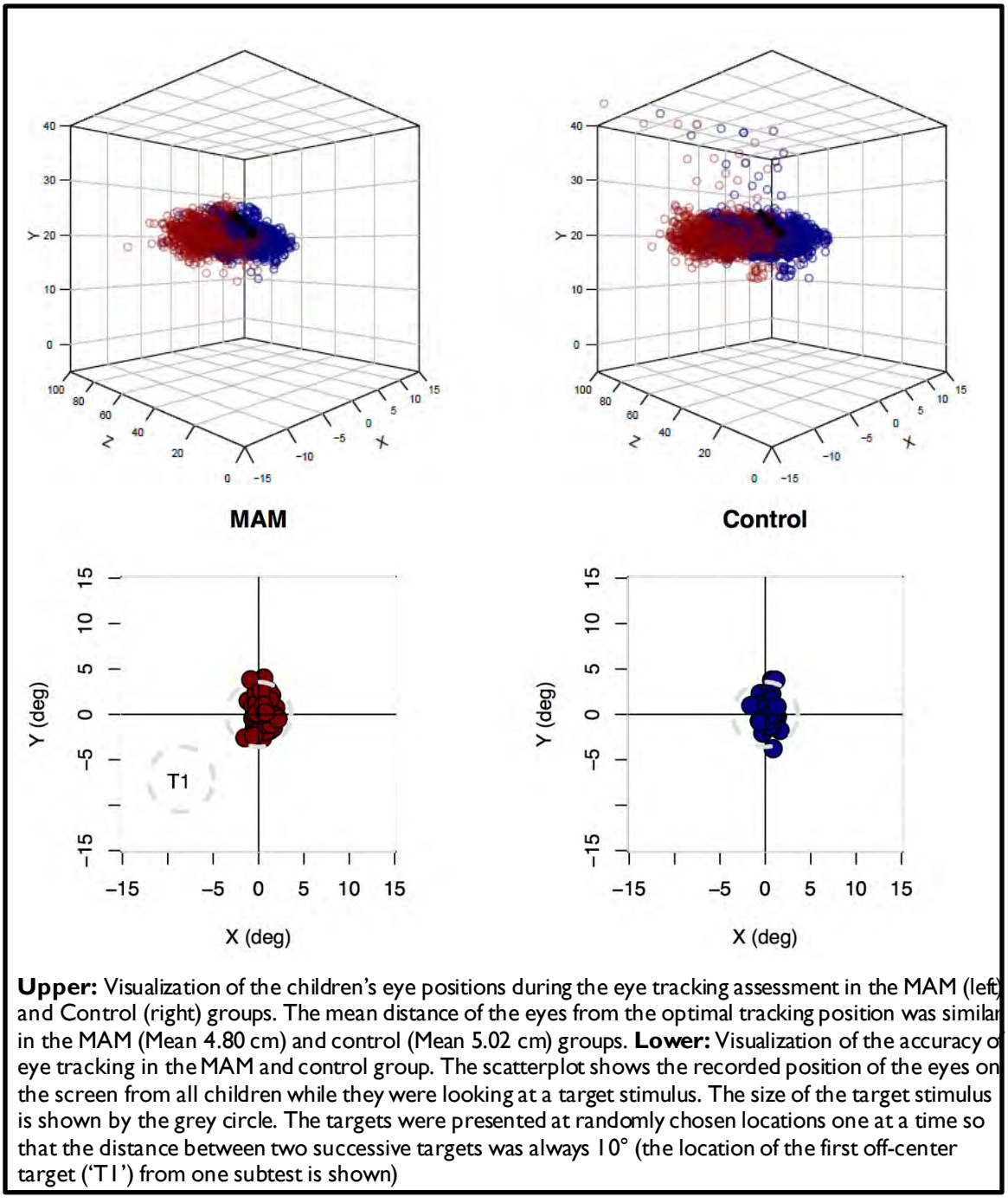
Raw data from the eye tracking tests of visual function were analyzed by using a library of computer scripts and functions (Leppänen et al., 2015). Key indicators of data quality were first extracted from the data, including measures of infant movement during testing, accuracy of recorded gaze positions, amount of data available for the estimation of the outcome variables and eye tracking noise (i.e., sample-to-sample variation in recorded gaze position). The main outcome variable from the eye tracking test, Saccadic Reaction Time (SRT), was defined as the time interval from the onset of the saccade target to the first entry of the gaze in the area of the saccade target (±1.4° margin). Secondary eye tracking outcomes were measures reflecting the spatial and temporal characteristics of visual fixations (i.e., periods when the eyes were stationary and fixated at a particular location of the image). For further detailed about the extraction of these variables from the raw eye tracking data, see **Appendix I**.

In statistical analysis of eye-tracking data, consistency of the eye-tracking outcomes within test sessions was assessed by calculating correlations between means extracted from odd and even-numbered test items (Pearson r) and across time (test-retest) by comparing means from Visit 1 and Visit 2. Odd-even split-half correlations have varied from .38 to .88 for measures of infant attention (Ahtola et al., 2014; Rose et al., 2012; Gillespie-Smith et al., 2016), and test-retest correlations from .53 to .76 over one to two-week interval and from .40 to .76 over a two-month interval in previous studies in high-resource settings (Cousijn et al., 2017; Leppänen et al., 2015).

The hypothesis that MAM is associated with neurocognitive slowing (i.e., longer SRTs) was tested by examining the effects of Group (MAM vs. Control) on mean log-transformed SRTs by using a linear mixed model (R & lme4 package). The Model included group, test block, trial number as fixed effects, child age and child gender as fixed effects, and a random intercept for each child as a random effect. To test the hypothesis that neurocognitive slowing in MAM resolves by supplementary feeding, the same model was fitted to the data with visit (pre vs. post) and Group x Visit as additional fixed effects. Corresponding models were fitted to the secondary indices of eye-tracking data (i.e., the coverage and duration of fixations while observing a social scene).

Data from the conventional tests of visual function were transformed into a binary outcome variable indicating the absence (0) or presence of the observed behavior (Partial/Complete, coded as 1) for each individual trial in each task, and the effects of Group (MAM vs. Control), Time (pre vs post), and test trial (1-4/5) on outcome behaviors (saccades, smooth pursuit and fixations) were modeled by using mixed effects logistic regression models, with random intercept for each child, as implemented in R and the lme4 package (R Core Team, 2014, Bates et al., 2015).

Figure 2. Children's Eye Positions (Upper) and Accuracy of Eye Tracking (Lower)



Upper: Visualization of the children's eye positions during the eye tracking assessment in the MAM (left) and Control (right) groups. The mean distance of the eyes from the optimal tracking position was similar in the MAM (Mean 4.80 cm) and control (Mean 5.02 cm) groups. **Lower:** Visualization of the accuracy of eye tracking in the MAM and control group. The scatterplot shows the recorded position of the eyes on the screen from all children while they were looking at a target stimulus. The size of the target stimulus is shown by the grey circle. The targets were presented at randomly chosen locations one at a time so that the distance between two successive targets was always 10° (the location of the first off-center target ('T1') from one subtest is shown)

5. Results

5.1 Eye Tracking Data Quality

Given that the current study used eye tracking in a novel setting, several variables were extracted to examine the quality of the measurements and the precision of the key outcome metrics (i.e., SRTs, fixation coverage and duration). Eye movements were recorded by remote cameras that allow for some degree of head movement during testing without data loss (**Figure 2**). Variability in head movement, calculated as root mean square of the distance from the eyes to the optimal position (x = 0 cm, y = 20 cm, and z = 56.4 cm) was 4.80 cm in the MAM group and 5.02 cm among controls, $p > .10$.

SRT Measures

For SRT analysis, the key indicators of data quality were error in the estimates of point of gaze, precision of the mean SRTs which is dependent on the number of valid trials and the within-subject consistency (reliability) of SRTs.

Mean error in the estimates of the point of gaze (i.e., distance from the returned value to the actual location of a calibration target) ranged from 0.68° to 1.14° (**Figure 2**). Given the focus on "object-level" attention shifts in the current analyses (i.e., detection of gaze shifts between two spatially-separate objects), the effect of this error (as well as the entire range of errors in the study) can be evaluated by contrasting the size of the error with the size of the saccade targets as well as the eccentricity of two successive targets. (This analysis, showing that all observed errors were within the boundaries of a single target object, is presented in **Figure 2.**)

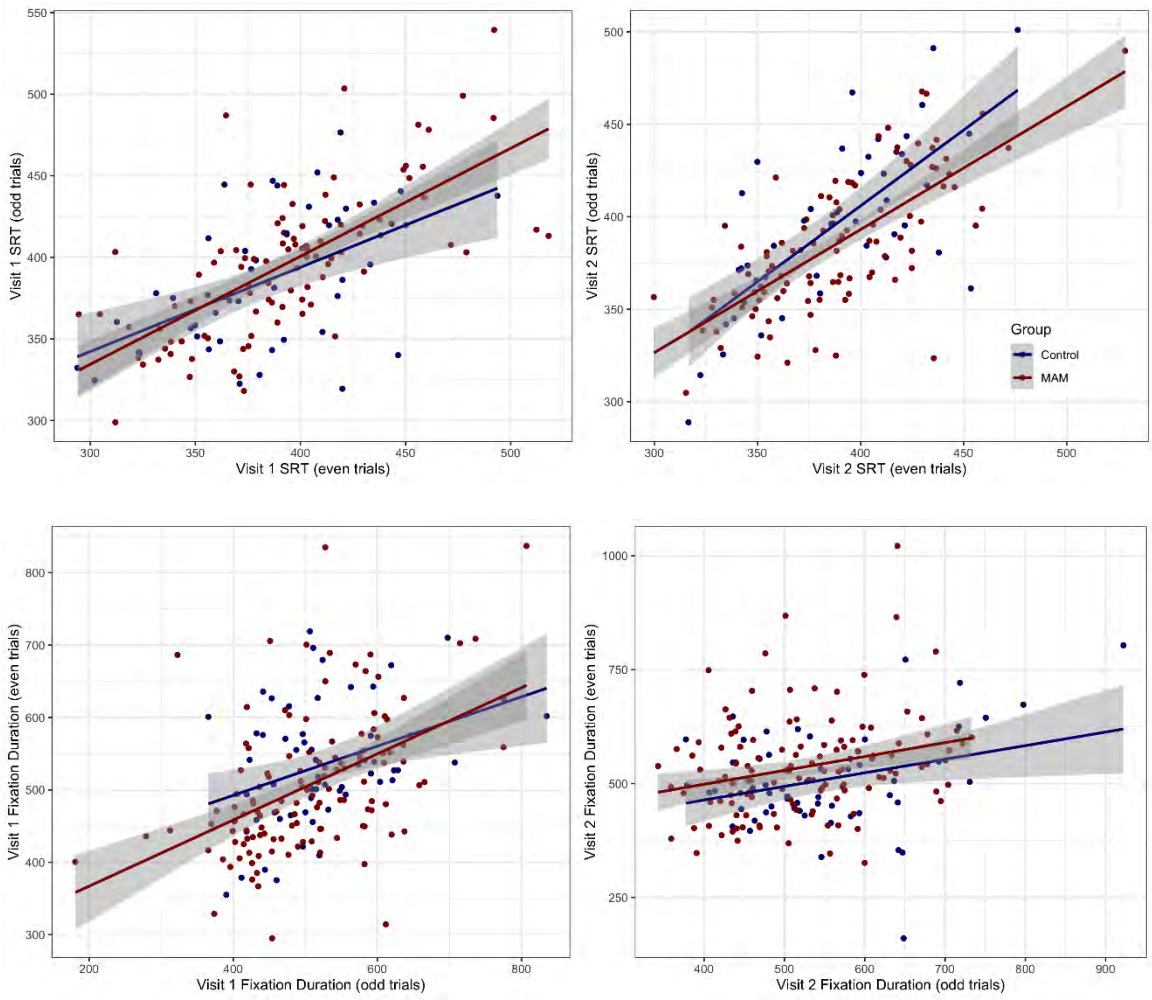
Precision (standard error) in the estimates of the mean SRT values, as obtained from SD/SQRT(n), was near optimal in the current analyses given that: i) SRTs based on <10 valid trials were not included in the analysis; ii) the mean number of valid trials ranged from 20 to 23; and iii) further increase in the number of valid trials would have led to negligible improvements in precision. The odd-even split-half reliability of the mean SRTs ranged from $r = .65$ to $r = .68$, and, the test-retest correlation (v1-v2) of mean SRT was $r = .59$ (**Figure 3**). None of these data quality indicators differed significantly between MAM and Controls, $ps > .10$.

Fixations

Spatiotemporal analyses of gaze fixations during visual exploration can be sensitive to noise, which is typically assessed by calculating the sample-to-sample change in xy position of the eyes during periods of stable fixations. Median noise levels was similar for MAM (V1 mean = 1.13, V2 mean = 1.19) and Control groups (V1 mean = 1.18, V2 mean = 1.10), $ps > 10$). Approximately 80 percent or more of the noise estimates were in the range that is considered optimal for the detection of the fixations with the algorithm used in the current study (<2°). Similar analyses were performed for the amount of valid data (in sec). The mean duration of valid data per scene did not differ significantly between MAM and Controls in the first visit (MAM 10.32 sec, Controls 10.25 sec), but MAM had more valid data in the second visit (MAM 11.31 sec, Controls 10.91 sec), $t(93.95) = 2.08$, $p = 0.04$. The odd-even split-half correlations were .26 (Visit

1) and .47 (Visit 2) for measures of the spatial coverage of fixations, and .41 and .46 for the mean duration of fixations. Test-retest correlations were .31 for spatial coverage and .44 for the mean duration of fixations (**Figure 3**).

Figure 3. Reliability of Saccadic Reaction Time (Upper) and Fixation Duration (Lower)



Reliability of Saccadic Reaction Time (upper) and Fixation Duration (lower) estimates in the baseline and post-treatment assessment. The reliability (or consistency) of the measurements within a testing session was estimated by examining the similarity of the means obtained when using data from odd (x-axis) and even-numbered (y-axis) test items in each assessment.

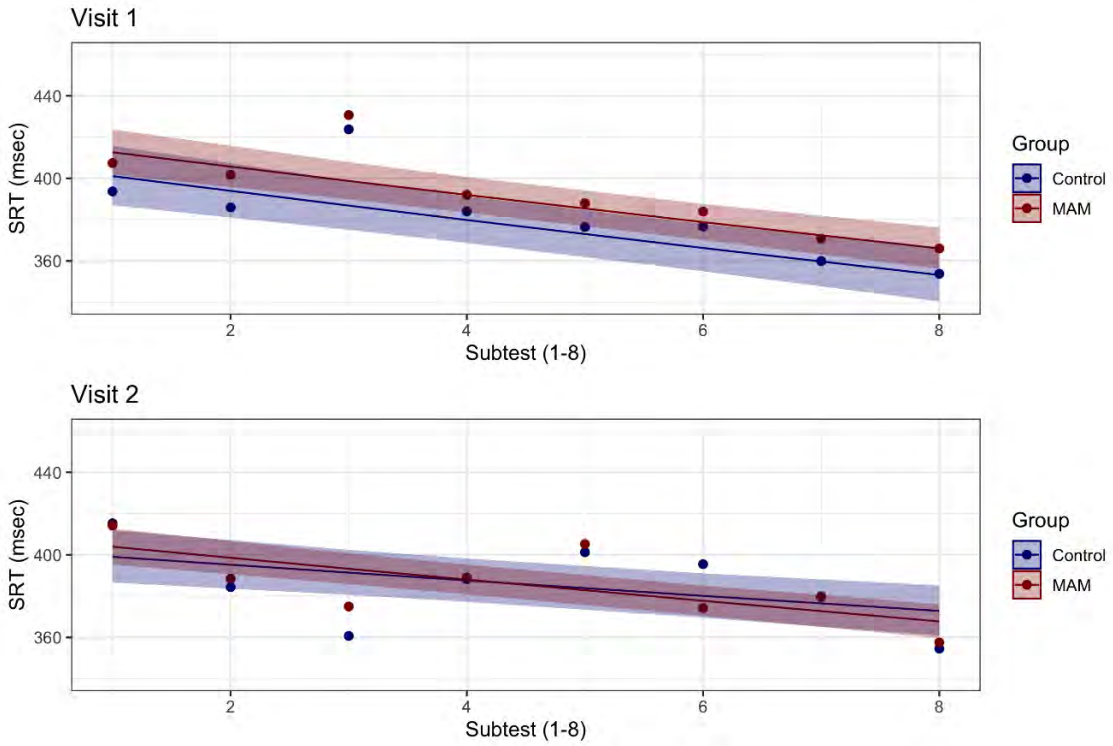
5.2 Effects of Supplementary Foods on Measures of Neurocognitive Function

Although the original aim of the study was to compare the four supplementary foods, the sample sizes were limited in some of the subgroups and for this reason, all statistical analyses were conducted by combining the MAM subgroups receiving different foods into a single MAM group.

SRTs

Response rates to the saccade targets on the computer screen were high for both groups, and there were no significant group differences in the rate of responding to the stimuli (Mean MAM_{visit1} = 95%, MAM_{visit2} = 94%, $Control_{visit1}$ = 94%, $Control_{visit2}$ = 95%). This result supports the feasibility of conducting the assessments in the two groups.

Figure 4. Mean Saccadic Reaction Times for Infants with MAM and Controls by Visit

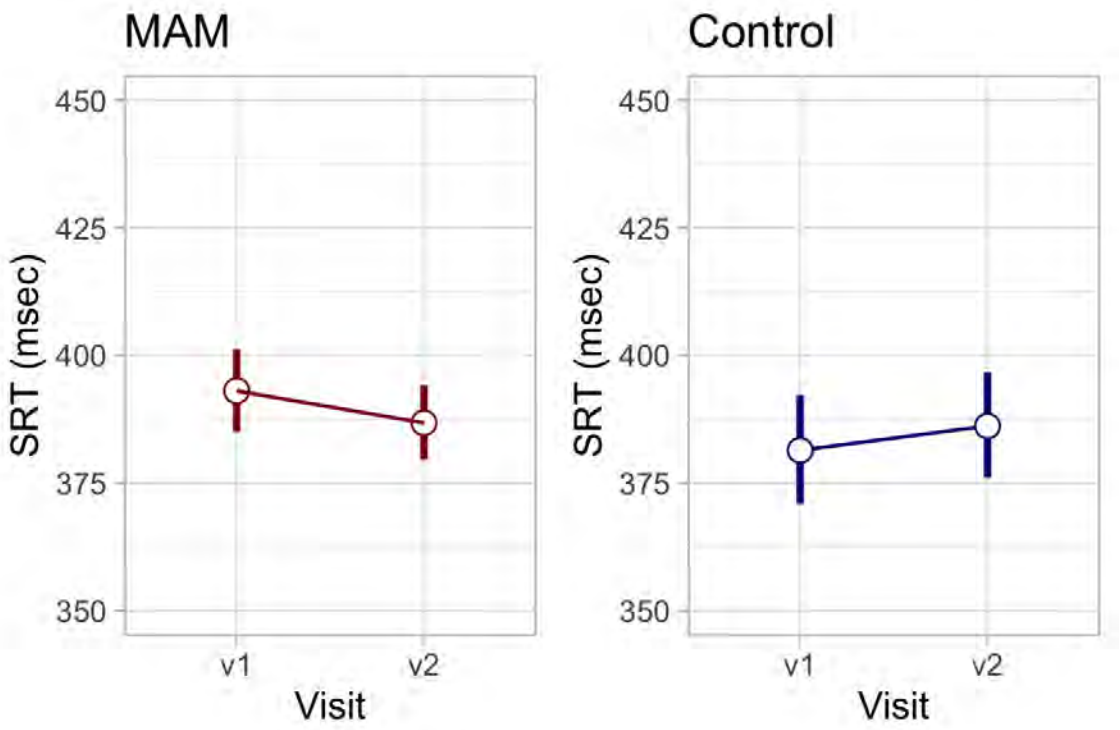


Mean saccadic reaction times (SRTs) for infants with MAM and Controls by visit. Data points are raw means for individual subtests with 5 saccade targets. Lines represent predicted effects of group and subtest number. Error bars represent 95% confidence intervals of the estimates.

The effects of Group (MAM vs. Control) on SRTs was examined by using data from the baseline assessment (i.e., before the supplementary feeding for MAM was started). MAM children had marginally slower SRTs than Controls in the baseline assessment (**Figure 3**), (χ^2 (1) = 2.93, p = .086). This difference was consistent across the subtests (blocks) in Visit 1 and independent of the general effect showing that the SRTs shortened over the course of the assessment, (χ^2 (1) = 194.45, p < .001 [**Figure 4**]). To examine neurocognitive recovery after supplementary feeding in MAM children, the MAM and the control groups were compared for the pattern of change in SRTs over time (pre vs. post). The Group by Time Interaction was significant, (χ^2 (1) = 7.14, p = .008). Further analyses of the change within each group showed that SRTs shortened 7 ms ±2 between

the first and the second visit in infants with MAM, (χ^2 (1) = 9.69, p = 0.002 [**Figure 5**]). No significant change in the mean SRT between the first and second visit was seen in the control group, p = .23 (**Figure 5**).

Figure 5. Change in Saccadic Reaction Times Between Visits at Baseline and Posttreatment



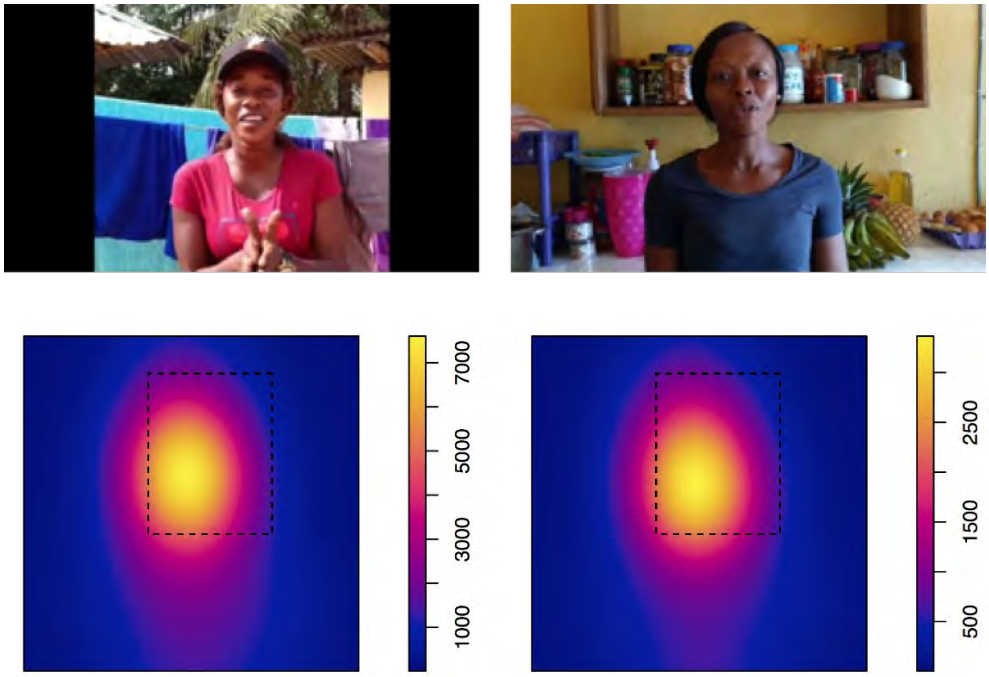
Change in saccadic reaction times (SRTs) between visits 1 (baseline) and 2 (posttreatment) in MAM and control children. Mean SRTs in are predicted values from a linear mixed effects model with group, visit, subtest number (1-8), saccade target (1-5), child age, and child gender as fixed factors. Error bars represent 95% confidence intervals of the point estimates.

Fixations

Spatial and temporal characteristics of fixations were examined by using data collected while the infants were viewing dyadic social scenes (four 10- to 15-second scenes per visit). This analysis showed that the local maxima of fixations correspond to the approximate location of the face in the stimulus videos (**Figure 6**), consistent with infants at this age typically exhibiting a strong perceptual predilection for faces[2]. There was no indication of group differences in the distribution of fixation at baseline or a differential pattern of changes in the distribution across time, p = .37, suggesting similar areal coverage of fixations for MAM and control children while viewing social scenes.

[2] As this study was not designed to examine the bias for face per se (e.g., faces were located in the same area of the display across the scenes), our data do not allow us to make a clear distinction between a face bias and other biases that may affect infant looking behavior (e.g., center bias), even though we consider the former as the most likely interpretation of the observed pattern.

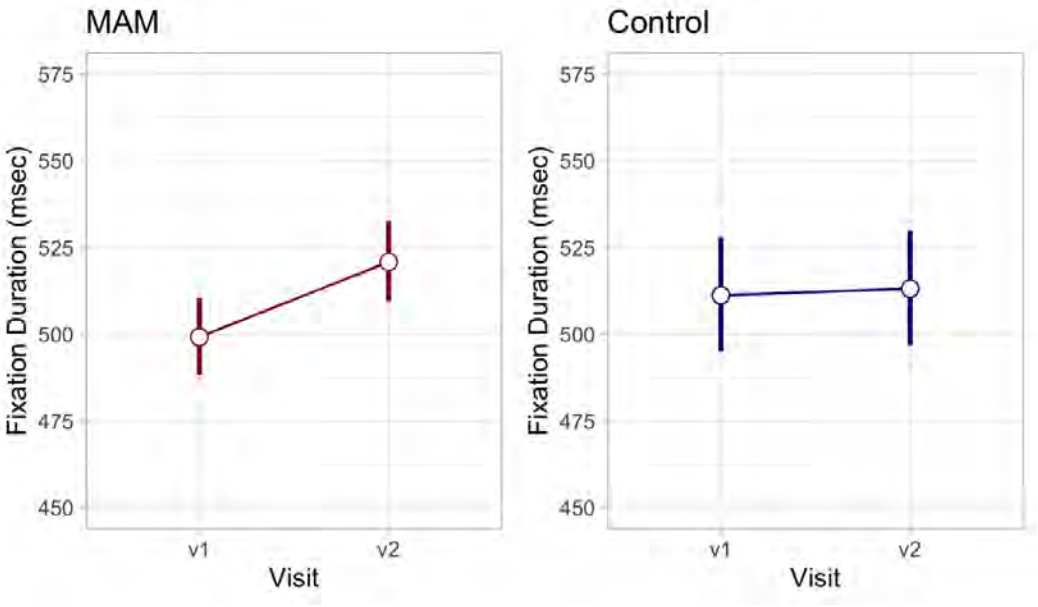
Figure 6. Social Scenes Excerpts (Upper) and Distribution of all Visual Fixations (Lower)



In the test assessing attention to social stimuli, infants were presented with short video recordings of Sierra Leonian women enacting a short story or a. Upper: Snapshots of two social scenes. Lower: Distribution of all fixations for the scenes (averaged across the four scenes presented at each visit). The dotted rectangle shows the area where the actor's face was on 75% of the frames in the video.

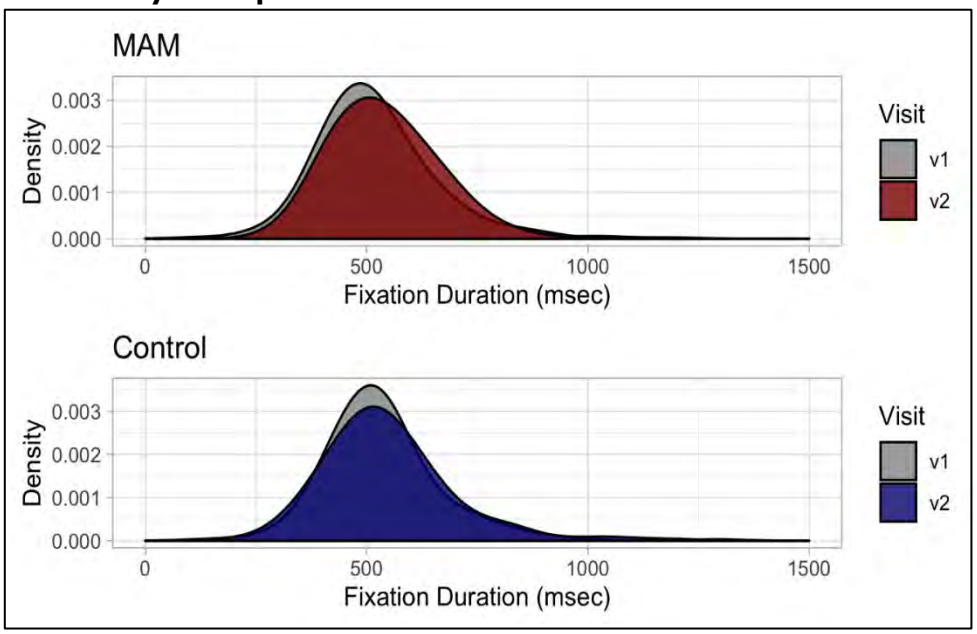
An established measure of the temporal characteristics of fixations in infants is the mean duration of fixations. The mean duration of fixations was calculated by averaging the duration of all fixations recorded while viewing the scenes (Mean number of fixations contributing to the average was 61.34 for Visit 1 and 66.55 for Visit 2). Relatively longer duration of fixations during free viewing in infancy has been associated with relatively better cognitive and behavioral control in mid childhood (Papageorgiou et al., 2014). The MAM and Control children exhibited a differential pattern of change in fixation duration over time, χ^2 (1) = 6.20, p = 0.01275 (corrected p = 0.032). Further analysis of the pattern of change within each group showed that the MAM group had marginally shorter fixations in the pre-test, χ^2 (1) = 3.95, p = 0.04680 (corrected p = 0.078), but the mean duration increased reliably for the MAM children in the post-treatment assessment, (χ^2 (1) = 33.31, p < .0001 (corrected p < .0001 [**Figure 7-8**]). No difference between visits was found in the control group, p = .43. The corrected p-values are corrected for false discovery rate by taking into account all the tests performed in the exploratory analysis of the fixation data.

Figure 7. Change in Fixation Duration Between Baseline and Posttreatment Visits in MAM and Control Children



Mean durations are predicted values from a linear mixed effects model with group, visit, scene (1-4), child age, and child gender as fixed factors. Error bars represent 95% confidence intervals for the point estimates.

Figure 8. Distribution Density Plots for Raw Fixation Durations for First and Second Visits by Group



Conventional Eye Movement Testing

The MAM and Control children did not differ in observational tests of eye movement control and these behaviors did not change in response to treatment, *p*s > .10. There was a significant effect of trial on saccades (χ^2 (1)=3.71, p = .05), smooth pursuit (χ^2 (1)=17.4, p < .001), and fixation (χ^2 (1) =70.51, p < .001), reflecting a general decline in the expression of the measured behaviors over the course of the test in both groups. This pattern did not vary by group, *p*s > .10. As a further analysis to verify the presence of the studied functions in each of the groups (i.e., saccades, fixation, and smooth pursuit), the number of infants in each group that showed no evidence for the targeted capacity were counted. This analysis was limited to data from Visit 1 (pretest) given the general decline in the targeted behaviors between the visits. There were no infants

without any partial/full saccades, smooth pursuit, or fixations in the MAM group and There were no infants without any partial/full saccades, smooth pursuit, or fixations in the MAM group and only one infant in the control group did show any response that would have been counted as a partial/complete response in the fixation condition.

6. Challenges and Limitations

The current study showed no major barriers for wider use of eye tracking in the field and in the assessment of short-term changes in neurocognitive function in groups of infants.

Supporting the technical feasibility of the methodology in this novel context, the key metrics of visual orienting and holding were obtained for most infants for two separate time points, showed expected levels of precision and within observer consistency. Completion rates for the tests were slightly lower than expected (76 percent, expected 81 percent for two successful tests). The team has made changes to the assessment technology that have helped to improve completion rates, and they are currently at the expected level. Some other aspects of the methodology can be further improved (e.g., spatial precision).

7. Conclusions

This report documents a novel use of eye tracking in the assessment of short-term changes in neurocognitive function in malnourished infants in settings that have very few facilities. However, they represent the circumstances of many resource-poor communities in sub-Saharan Africa.

Consistent with the predicted effect of malnutrition on neurocognitive function, infants with MAM had marginally slower saccadic reaction times and shorter fixations than Controls did. The small size of these differences may reflect the fact that while the two groups differed in nutritional status, the participants were all children from environments that are disadvantaged by many measures. When compared to values reported in the literature, the mean SRTs in the current study were clearly slower than the mean SRTs typically found for infants in this age range. Direct comparisons are

complicated by procedural differences between studies, but the mean SRTs were 400 ms for MAM and 390 for Controls in the current study. The corresponding values in previous studies in Western populations have been consistently shorter, ranging from 270 to 375 msec (Caudill et al., 2017, Cousijn et al., 2017; Rose et al., 2002; Wass & Smith , 2014). Thus, our findings are consistent with the hypothesis that MAM has neurocognitive effects, but they also raise the possibility that the effects of MAM on neurocognition may not stand out as a strong effect in resource-poor settings, possibly reflecting the fact that most children are affected by various poverty-related risk factors in these environments (McCoy et al., 2016).

As a first direct demonstration of neurocognitive benefits of supplementary feeding, our results showed a significant change towards improved visual orienting (saccadic reaction time) and attentional focusing (mean duration of fixations) in the MAM group after four weeks of supplementary feeding. The interpretation of these changes as genuine improvements in neurocognitive function is supported by the consistency of the observed pattern across the two measures, by previous data showing that specific nutritional supplements result in similar changes in saccadic reaction time in high-resource settings in the U.S. (Caudill et al., 2017), by the established status of faster orienting and longer holding as markers of cognitive function (Dougherty & Haith, 1997; Papageourgiou et al., 2014), and, lastly, by the apparent functional benefits that come from faster orienting and longer durations of stable fixations during visual exploration. There were also no indications in the data that the differential change in the MAM and control group in either of the two measures reflected secondary effects arising from the test administration becoming more feasible between the two visits. The possibility remains, however, that the size of the improvements in the MAM group may change by supplement type.

To conclude, the results of this study demonstrate the promise of eye tracking in the assessment of neurocognitive status in malnourished children in low-resource field settings. The results also show novel evidence for the functional benefits of supplementary feeding of children with MAM.

In the future, this approach could be used to assess the effectiveness of different food compositions in the treatment of undernutrition in food-insecure environments. First, the current method can be optimized for larger scale use as a standardized, widely-applicable marker of neurocognitive outcomes in children in different environments and of varying age. Second, future studies will be needed to examine whether the treatment effects that were visible within a relatively short time scale in the current study (i.e., after four weeks of treatment) are maintained or increased over a longer period of time, and whether they transfer to more complex cognitive skills later in childhood.

Given the hierarchical nature of brain development and skill formation, even small improvements in elementary sensory and attentional functions in the early stages of development may result in large secondary effects on more advanced neurocognitive processes and, thereby, have a long-term, accumulating impact on skill formation. Finally,

the sample size in the current sub-study was limited for reliable analysis of differences among the four foods, but future studies with larger samples can use the current technology as an automated, standardized, and scalable method to compare neurocognitive outcomes across different food compositions and interventions. Together, these steps will be significant in finding the most effective interventions for early childhood malnutrition, and in helping the field to move beyond anthropometric outcomes to other outcomes that are important in the development of human capacity and well-being.

Acknowledgements

This study would not be possible without the support of the USAID Bureau for Humanitarian Assistance (BHA) and the legacy Office of Food for Peace (FFP) and their ongoing commitment to improving Title II programming in order to address food insecurity in vulnerable populations. The time and support offered to the Food Aid Quality Review by Title II awardees, both headquarters and field staff, as well as Food for Peace Officers abroad and in Washington, D.C. and the Policy and Technical Division of FFP headquarters were also invaluable to informing this study.

References

Ahtola E, Stjerna S, Yrttiaho S, et al. (2014). Dynamic eye tracking based metrics for infant gaze patterns in the face-distractor competition paradigm. *PLoS One*, 5, e97299.

Alahyane N, Lemoine-Lardennois C, Tailhefer C, et al. (2016). Development and learning of saccadic eye movements in 7- to 42-month-old children. *Journal of Vision, 1*, 1-12.

Bargary G, Bosten JM, Goodbourn PT, et al. (2017). Individual differences in human eye movements: an oculomotor signature? *Vision Research*, 141, 157-169.

Bates D, Mächler, M, Bolker B, et al. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1.

Birch EE, Carlson SE, Hoffman DR, et al. (2010). The DIAMOND (DHA Intake And Measurement Of Neural Development) study: a double-masked, randomized controlled clinical trial of the maturation of infant visual acuity as a function of the dietary level of docosahexaenoic acid. *American Journal of Clinical Nutrition*, 9, 848-859.

Brenna JT. (2011). Animal studies of the functional consequences of suboptimal polyunsaturated fatty acid status during pregnancy, lactation and early post-natal life. *Maternal and Child Nutrition*, 2, 59-79.

Carlson SE, Werkman SH, Rhodes PG, et al. (1993). Visual-acuity development in healthy preterm infants: effect of marine-oil supplementation. *American Journal of Clinical Nutrition*, 58, 35-42.

Caudill MA, Strupp BJ, Muscalu L, et al. (2017). Maternal choline supplementation during the third trimester of pregnancy improves infant information processing speed: a randomized, double-blind, controlled feeding study. *FASEB Journal*, 32, 2172-2180.

Colombo J, Carlson SE, Cheatham CL, et al. (2011). Long-chain polyunsaturated fatty acid supplementation in infancy reduces heart rate and positively affects distribution of attention. *Pediatric Research*, 70, 406-410.

Colombo J, Carlson SE, Cheatham CL, et al. (2013). Long-term effects of LCPUFA supplementation on childhood cognitive outcomes. *American Journal of Clinical Nutrition*, 98, 403-412.

Colombo J, Gustafson KM, Gajewski BJ, et al. (2016). Prenatal DHA supplementation and infant attention. *Pediatric Research*, 80, 656-662.

Constantino JN, Kennon-McGill S, Weichselbaum C, et al. (2017). Infant viewing of social scenes is under genetic control and is atypical in autism. *Nature*, 547, 340-344.

Cousijn J, Hessels, RS, Van der Stigchel S, et al. (2017). Evaluation of the psychometric properties of the gap-overlap task in 10-month-old infants. Infancy, 22, 571-579.

Diau GY, Hsieh AT, Sarkadi-Nagy EA, et al. (2005). The influence of long chain polyunsaturate supplementation on docosahexaenoic acid and arachidonic acid in baboon neonate central nervous system. *BMC Medicine*, June 23, 3-11.

Dougherty TM, Haith MM. (1997). Infant expectations and reaction time as predictors of childhood speed of processing and IQ. *Developmental Psychology*, 1, 146-155.

El-khayat H, Shaaban S, Emam EK, et al. (2007). Cognitive functions in protein-energy malnutrition: in relation to long chain-polyunsaturated fatty acids. *Pakistanian Journal of Biological Sciences*, 10, 1773-1781.

Forssman L, Ashorn P, Ashorn U, et al. (2017). Eye-tracking-based assessment of cognitive function in low-resource settings. *Archives of Diseases in Childhood*, 4, 301-302.

Galler JR, Bryce CP, Zichlin ML, et al. (2013). Malnutrition in the first year of life and personality at age 40. *Journal of Child Psychology and Psychiatry*. 54, 911-919.

Galler JR, Ramsey F. (1989). A follow-up study of the influence of early malnutrition on development: behavior at home and at school. *Journal of the American Academy of Child and Adolescent Psychiatry*, 28, 254-261.

Galler JR, Ramsey F, Solimano G, et al. (1983). The influence of early malnutrition on subsequent behavioral development: II. Classroom behavior. *Journal of the American Academy of Child and Adolescent Psychiatry*, 22, 16-22.

Georgieff MK. (2007). Nutrition and the developing brain: nutrient priorities and measurement. *American Journal of Clinical Nutrition*, 85, 614S-620S.

Gillespie-Smith K, Boardman JP, Murray IC, et al. (2016). Multiple measures of fixation on social content in infancy: evidence for a single social cognitive construct? *Infancy*, 2, 241-257.

Gould JF, Makrides M, Colombo J, et al. (2014). Randomized controlled trial of maternal omega-3 long-chain PUFA supplementation during pregnancy and early childhood development of attention, working memory, and inhibitory control. *American Journal of Clinical Nutrition*, 99, 851-859.

Hernik M, Broesch T. (2018). Infant gaze following depends on communicative signals: an eye-tracking study of 5- to 7-month-olds in Vanuatu. *Developmental Science*, 22, e12779.

Hessels RS, Niehorster DC, Kemner C, et al. (2017). Noise-robust fixation detection in eye movement data: identification by two-means clustering (I2MC). *Behavior Research Methods*, 5, 1802-1823.

Hsieh JC, Liu L, Zeilani M, et al. (2015). High-oleic ready-to-use therapeutic food maintains docosahexaenoic acid status in severe malnutrition. *Journal of Pediatric Gastroenterology & Nutrition*, 1, 138-143.

Irving EL, González EG, Lillakas L, Wareham J, et al. (2011). Effect of stimulus type on the eye movements of children. *Investigative Ophthalmology & Visual Science*, 2, 658-664.

Itti, L. (2015). New eye-tracking techniques may revolutionize mental health screening. *Neuron*, 3, 442-444.

Jones PR, Kalwarowsky S, Atkinson J, et al. (2014). Automated measurement of resolution acuity in infants using remote eye-tracking. *Investigative Ophthalmology & Visual Science*, 12, 8102-8110.

Jones W, Klin A. (2013). Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism. *Nature*, 7480, 427-431.

Kulke L, Atkinson J, Braddick O. (2015). Automatic detection of attention shifts in infancy: eye tracking in the fixation shift paradigm. *PLoS ONE* ,12, e0142505.

Leppänen JM, Forssman L, Kaatiala J, et al. (2015). Widely applicable MATLAB routines for automated analysis of saccadic reaction times. *Behavior Research Methods*, 2, 538-548.

Lu C, Black MM, Richter LM. (2016). Risk of poor development in young children in low-income and middle-income countries: an estimation and analysis at the global, regional, and country level. *Lancet Global Health*, 4, e916-e922.

Makrides M, Gibson RA, McPhee, et al. (2010). Effect of DHA supplementation during pregnancy on maternal depression and neurodevelopment of young children: a randomized controlled trial. *JAMA*, 304, 1675-1683.

McCoy DC, Peet ED, Ezzati M, et al. (2016). Early childhood developmental status in low- and middle-income countries: national, regional, and global prevalence estimates using predictive modeling. *PLoS Medicine*, 6, e1002034.

Morgane PJ, Austin-LaFrance R, Bronzino J, et al. (1993). Prenatal malnutrition and development of the brain. *Neuroscience and Biobehavior Reviews*, 17, 91-128.

Munoz DP, Everling S. (2004). Look away: the anti-saccade task and the voluntary control of eye movement. *Nature Reviews Neuroscience*, 3, 218-228.

Nelson CA III, Zeanah CH, Fox NA, et al. (2007). Cognitive recovery in socially deprived young children: the Bucharest early intervention project. *Science*, 318, 1937-1940.

Nishimura T, Takei N, Tsuchiya KJ, et al. (2016). Identification of neurodevelopmental trajectories in infancy and of risk factors affecting deviant development: a longitudinal birth cohort study. *International Journal of Epidemiology*, 45, 543-553.

Olsson M, Teär Fahnehjelm K, Rydberg A, et al. (2015). Ocular Motor Score (OMS): a clinical tool to evaluating ocular motor functions in children. Intrarater and inter-rater agreement. *Acta Ophthalmology*, 5, 444-449.

Papageorgiou KA, Smith, TJ, Wu R, et al. (2014). Individual differences in infant fixation duration relate to attention and behavioral control in childhood. *Psychological Science*, 7, 1371-1379.

Pearson BC, Armitage KR, Horner CW, et al. (2007). Saccadometry: the possible application of latency distribution measurement for monitoring concussion. *British Journal of Sports Medicine*, 41, 610-612.

Peirce JW. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, 1-2, 8-13.

Peter CJ, Fischer LK, Kundakovic M, et al. (2016). DNA Methylation Signatures of Early Childhood Malnutrition Associated With Impairments in Attention and Cognition. *Biological Psychiatry*, 10, 765-774.

Pyykkö J, Forssman L, Maleta K, et al. (2019). Early development of visual attention in infants in rural Malawi. *Developmental Science*, 5, e12761, doi: 10.1111/desc.12761.

R Core Team. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/

Rose SA, Feldman JF, Jankowski JJ. (2003). The building blocks of cognition. *Journal of Pediatrics*, 4, S54-S61.

Rose SA, Feldman JF, Jankowski JJ. (2012). Implications of infant cognition for executive functions at age 11. *Psychological Science*, 11, 1345-1355.

Rose SA, Feldman JF, Jankowski JJ, et al. (2002). A longitudinal study of visual expectation and reaction time in the first year of life. Child Development, 1, 47-61.

Suri DJ, Rosenberg IH. (2018). Beyond Z-scores: Measures to Advance Prevention and Treatment Outcomes in Child Malnutrition. Summary of Proceedings from a Symposium at the International Congress of Nutrition, Buenos Aires, Argentina on October 16, 2017. Report to USAID. Boston, MA: Tufts University

Tseng PH, Cameron IG, Pari G, et al. (2013). High-throughput classification of clinical populations from natural viewing eye movements. *Journal of Neurology*, 1, 275-284.

Walker SP, Wachs TD, Grantham-McGregor S, et al. (2011). Inequality in early childhood: risk and protective factors for early child development. *Lancet*, 9799, 1325-1338.

Wass SV, Smith TJ. (2014). Individual differences in infant oculomotor behavior during the viewing of complex naturalistic scenes. *Infancy*, 4, 352-384.

Westerberg AC, Schei R, Henriksen C, Smith L, et al. (2011). Attention among very low birth weight infants following early supplementation with docosahexaenoic and arachidonic acid. *Acta Paediatrica*, 1, 47-52.

Ylitalo, A. (2017). Statistical inference for eye movement sequences using spatial and spatio-temporal point processes. University of Jyväskylä, Department of Mathematics and Statistics. http://urn.fi/URN:ISBN:978-951-39-7064-2.

Appendix I: Procedure and Stimuli in the Eye-Tracking Test

The eye tracking test consisted of two sessions that lasted three to four minutes. In each session, the infant saw alternating blocks of dynamic stimuli, including: a) 3-9 calibration targets; b) 4x5 saccade targets; c) 2 videos depicting a dyadic social scene; and d) 2 filler videos. The order of the blocks was always the same: 1) calibration targets; 2) social scene 1; 3) saccade targets 1-5; 4) filler 1; 5) saccade targets 6-10; 6) social scene 2; 7) saccade targets 11-15 ;8) filler 2; and 9) saccade targets 16-20. Infants completing both test sessions saw a total of six to eighteen calibration targets, eight blocks of five saccade targets (forty saccade targets in total), and four social scenes on each visit.

Calibration Targets

Calibration targets were short sequences of stimuli, consisting of an audiovisual animation (5.7° x 5.7°, duration: 1000-3000 msec), an isoluminant white disc (1.3°x 1.3°, 1000 ms), and reappearance of the animation (5.7° x 5.7°, 1000 msec). The sequences were presented in the screen center (0°,0°), bottom-left (-11.4°, -11.4°) and top right (11.4°, 11.4°). The targets were presented up to two times if less than 75 percent of the recorded point of gaze samples were valid for more than one of the targets. Samples recorded during the white discs that fell in a 9°x 9° rectangle surrounding the disc and a similarity transformation estimator (https://pypi.org/project/nudged/) were used to calibrate the raw point of gaze estimates from the eye tracker. The accuracy of the calibrated point-of-gaze estimates was verified by calculating the Euclidean distance of the corrected estimates from the location of the targets at the start of the experiment and in half-way of the experiment. Similar comparisons were performed for the raw uncalibrated point of gaze estimates. The method that provided the lowest error was used. Corrected values were superior for 88 percent of infants for Visit 1 data and 83 percent for Visit 2 data.

Saccade Targets

Saccade targets were colorful cartoon animations of common objects (e.g., fish, bird, pig, rabbit or a human face), subtending a 5.7° x 5.7° visual angle. The targets were presented one at a time without onset delay so that the first target was always presented in the center of the screen (0°,0°) and each subsequent target was presented in a new, randomly-chosen on-screen location 10° away. After presented on screen, the animations remained paused at the first frame until the infant looked at the animation (i.e., the first gaze point overlapping with the target area was recorded) or a 1000-msec wait period elapsed. The animation was subsequently played for 2500 msec. After the initial stimulus in the center of the screen (not used in the analysis given that the starting position for the first saccade was unknown), Infants saw a total of five saccade targets in one block and total of eight blocks on each visit. The decision to use clearly discernible colorful animations, randomly-chosen target locations and novel targets for each presentation instead of more uniform visual stimuli and fixed locations (Bargary et al., 2017) was based on infants' known proclivity for novelty and on evidence showing

superior response rates to colorful, pictorial stimuli in young children (Irving et al., 2011)

Filler Items

To reduce monotony, filler items were presented in between the blocks of saccade targets. The items were short (five second) video recordings of two Sierra Leonian, Mende-speaking female models who greeted the child with a friendly tone of voice. The videos were taken in various environments (i.e., in an outdoor yard or in a living room area) and depicted two of the same models used in the dynamic social scene stimuli (described below).

Social Scenes

Similar to previous studies (Constantino et al., 2017; Jones & Klin, 2013) social scenes were 9.5 to 15-sec video recordings of Sierra Leonian female models who greeted the child and continued by enacting a short story or a song appropriate for the participants' age range. The videos were taken in various environments (i.e., in a kitchen, in a yard, in a living room area). Infants saw a total of four scenes on each visit.