A MULTI-COUNTRY, MULTI-CENTRE, THREE-ARM, PARALLEL GROUP, OPEN LABEL, RANDOMIZED TRIAL OF THREE SUPPLEMENTS FOR THE NUTRITIONAL MANAGEMENT OF MODERATE WASTING IN CHILDREN 6-24M OF AGE WITH ACUTE ILLNESS

STATISTICAL ANALYSIS PLAN

Version Number: 1.4

Date: October 7, 2025

Full Title	Effectiveness of a microbiome-directed food to promote programmatic and sustained nutritional recovery among children with uncomplicated acute malnutrition
Trial Registration	ISRCTN53213318
Sponsor	Bill & Melinda Gates Foundation
Primary Contact	Ayesha De Costa, World Health Organization
Study Statistician	Christopher Sudfeld, Harvard T.H. Chan School of Public Health
Primary Contact Signature and Date	
Study Statistician Signature and Date	 October 20, 2025

Contents

1 Version History

Version number	Version date	Prepared by	Description of the SAP revision
1.0	November 3, 2022	Sudfeld	-First draft of SAP
1.1	January 9, 2022	Sudfeld	-Defined secondary outcomes for second randomization and exploratory outcomes for combined first and second randomization analyses
1.2	June 7, 2023	Sudfeld	-Clarified to 'real-time' second randomization and that pilot participants will be excluded from trial analyses.
1.3	May 26, 2025	Sudfeld	-Updated secondary outcomes (removed proportion ΔWLZ $\geq$0.5; added proportion SAM) -Refined secondary outcomes measured at 12 weeks and 24 weeks -Defined Per Protocol (PP) analyses -Specified exploratory analyses for effect modifiers
1.4	October 7, 2025	Sudfeld	-Clarified the tricep as measure for secondary skinfold thickness outcome -Refined definition of three adherence / consumption measures -Specified 'any' hospitalization as binomial secondary outcome (non-repeated)

2 Introduction

2.1 Background and Rationale

There is limited evidence on the management of moderate malnutrition (MAM) among children. While there is clear guidance on the management of children with severe wasting, there is currently no consistent guidance on how best to manage children with moderate wasting. While there is no consistent guidance on the management of moderate wasting in children and more evidence is needed, there is increasing recognition that treatment should be offered to those at the highest risk of death and deterioration, as children with moderate wasting and additional risk factors (e.g., infections) have higher risks of morbidity and mortality.

Moderate wasting is associated with a 3-fold higher risk of mortality in children under 5 years of age compared with their well-nourished peers. These deaths largely occur from infectious illnesses. The interactions between episodic infections and malnutrition are complex and bidirectional. Recent studies have elucidated possible mechanisms by which malnourished children are rendered more vulnerable to infection. And how an infection in turn worsens moderate wasting, resulting in deterioration to severe wasting and ultimately death. Acute infectious illness is a clear risk factor increasing the risk of adverse outcomes in moderately wasted children.

Despite mechanistic work in this area, there are no clinical trials that have been done in this high-risk subgroup to investigate if it is possible to break the malnutrition-infection cycle and decrease the risk of deterioration/death by initiating interventions targeting the nutritional (anthropometric) deficit. The NUTRIMAM trial proposes to evaluate different nutritional supplements in young children who are moderately wasted and present with an acute infection. The current IMCI guidelines on the feeding of moderately wasted children presenting to a health facility with illness recommend counseling on increased feeding using home-available foods. The Essential Nutrition Actions document (2013) recommends the provision of an increased amount of food after an acute illness. The findings from the NUTRIMAM trial are expected to fill in the gaps concerning guidance on the feeding of moderately wasted children with acute illness.

2.2 Objectives

The NUTRI-MAM trial will aim to evaluate the efficacy of locally available foods (LAF) or microbiota-directed supplementary food (MDSF) compared to ready-to-use supplemental foods (RUSF) when given to moderately wasted children presenting to health facilities with an acute illness, on initial anthropometric recovery (within 12 weeks of enrolment) and sustained recovery (at 24 weeks after enrolment).

Objectives

1. To determine the effect of (a) Locally Available Foods (LAF) compared to Ready to Use Supplementary Food (RUSF) and (b) microbiota-directed supplementary food (MDSF) compared to RUSF when given to moderately wasted children aged 6-24 months presenting to a health facility with an acute illness, on anthropometric recovery within 12 weeks of enrolment.

2. To determine the effects of (a) LAF compared to RUSF and (b) MDSF compared to RUSF, on sustained anthropometric recovery at 24 weeks after enrolment.

3. To determine the effect of a sustained recovery intervention compared with standard care (national standard of care following recovery) on sustained anthropometric recovery at 24 weeks after enrolment in the study.

3 Study Methods

3.1 Trial Design

NUTRI-MAM is a multi-center, open-label, sequential multiple assignment randomized trial that will be implemented in 5 countries - three in Asia (Bangladesh, India, and Pakistan) and two in Africa (Mali and Tanzania). The trial is a *Sequential Multiple Assignment Randomized Trial (SMART)* with two randomization stages. The first randomization stage occurs at enrollment and during 'Phase 1'. The second randomization stage occurs after 12 weeks and is referred to as 'Phase 2'. In Phase 2, only children who reached the anthropometric recovery in Phase 1 ("responders") will be randomized.

Phase 1 (Stage 1 interventions) of the study is focused on achieving anthropometric recovery. The first phase of the trial tests the effect of three types of food (RUSF, LAF, and MDSF) for the management of moderate wasting (WLZ <-2 but ≥ -3, or MUAC <125 mm but ≥115mm) on anthropometric recovery within 12 weeks of enrolment. For the study, anthropometric recovery is defined as WLZ > -2.0 or MUAC ≥ 125 mm. Children will be randomized to receive either RUSF, LAF, or MDSF. Weight, length, and MUAC will be measured for all enrolled children every week during this phase to determine if anthropometric recovery has occurred.

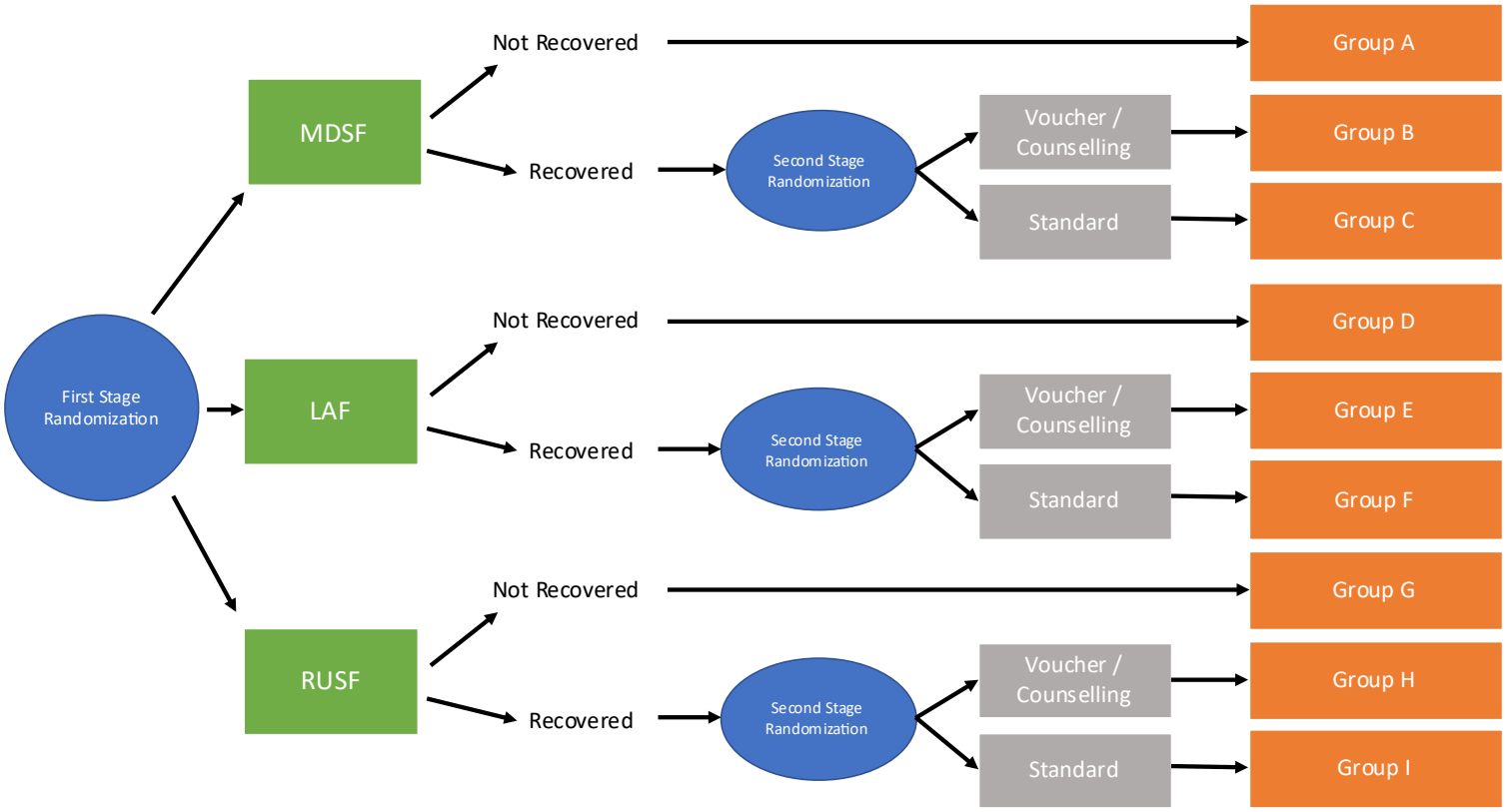
Phase 2 (Stage 2 interventions) of the study is focused on sustaining anthropometric recovery. Children who recover from moderate wasting in phase 1 ("responders") will be randomized a second time to receive either follow-up intervention or standard care (the national standard of care for recovered children in the country). Caregivers of children randomized to the follow-up intervention group will receive a sustained recovery package or 12 weeks after the second randomization (i.e., 24 weeks after initial enrolment into phase 1) and continued counseling (once a month). The standard care group will receive no subsequent additional intervention in phase 2 from week 12 to week 24 after initial enrolment. At the end of the follow-up period, 24 weeks after enrolment, weight, length, and MUAC will be measured in all children randomized in phase 2 to assess if recovery has been sustained.

3.2 Randomization

Randomization will occur in two stages and will be stratified by study site. The first stage is for the MDSF, LAF, and RUSF interventions, and then the second randomization to the sustained recovery intervention package or standard of care will only occur among children who have anthropometric recovery by 12 weeks. We will use 'real-time' randomization for the second randomization. In the up-front approach, it is important to note that those do not recover will not receive the second-stage randomization intervention. Figure 1 below illustrates the

randomization points and then the analysis groups. Note there are 6 groups: MDSF + sustained recovery intervention (A+B), MDSF + standard (A+C), LAF + sustained recovery intervention (D+E), LAF + standard (D+F), RUSF + sustained recovery intervention (G+H), RUSF+ standard (G+I)

Figure 1. Nutrimam SMART flow and analysis groups.



First-Stage Randomization Interventions: MDSF vs LAF vs RUSF

Primary Outcomes

1) **12-week Anthropometric Recovery**: The outcome will be defined by reaching the anthropometric recovery at any visit time point up to 12 weeks. Anthropometric recovery will be defined as reaching a WLZ ≥ -2.0 or MUAC ≥ 125 mm based on each participant's enrollment WLZ and MUAC. If a child was enrolled in the trial due to WLZ criteria alone (WLZ <-2 but ≥ -3), then recovery will be defined by reaching a WLZ ≥ -2.0. If a child was enrolled due to MUAC criteria alone (MUAC <125 mm but ≥15mm), then recovery will be defined by reaching a MUAC ≥ 125 mm. If a child was enrolled due to both WLZ and MUAC criteria, then recovery is defined as the child reaching either WLZ ≥ -2.0 or MUAC ≥ 125 mm.

2) **24-week Sustained Anthropometric Recovery**: The 24-week sustained anthropometric recovery outcome will be considered a primary outcome for LAF vs RUSF and MDSF vs RUSF. 24-week sustained recovery will be considered a secondary outcome for the LAF vs MDSF comparison.

   Sustained anthropometric recovery will be defined as meeting the 12-week anthropometric outcome by 12 weeks and continuing to have anthropometric recovery at 24 weeks. If a child was enrolled in the trial due to WLZ criteria alone (WLZ <-2 but ≥ -3), then sustained recovery will be defined as having a WLZ ≥ -2.0 by 12 weeks to meet the 12-week anthropometric recovery outcome and a WLZ ≥ -2.0 at the 24-week visit. If a child was enrolled due to MUAC criteria alone (MUAC <125 mm but ≥115mm), then sustained recovery will be defined by having a MUAC ≥ 125 mm by 12 weeks for the 12-week anthropometric recovery outcome and a MUAC ≥ 125 mm at the 24-week visit. If a child was enrolled due to both WLZ and MUAC criteria, then sustained recovery is defined as the child having either WLZ ≥ -2.0 or MUAC ≥ 125 mm by 12 weeks for the 12-week anthropometric recovery outcome and having either WLZ ≥ -2.0 or MUAC ≥ 125 mm at the 24-week visit.

Secondary Outcomes

1) Time to anthropometric recovery within the first 12 weeks (time-to-event): This will be defined as the number of days to anthropometric recovery within the first 12 weeks. The day of anthropometric recovery will be the study visit day when recovery was recorded.
2) Mean change in MUAC from (i) enrollment to 12 weeks and (ii) enrollment to 24 weeks: These two outcomes will be defined as the 12-week MUAC value minus the baseline MUAC value and the 24-week MUAC value minus the baseline MUAC value.
3) Mean change in WLZ from (i) enrollment to 12 weeks and (ii) enrollment to 24 weeks: These two outcomes will be defined as the 12-week WLZ value minus the baseline WLZ value and the 24-week WLZ value minus the baseline WLZ value.
4) Mean change in WAZ from (i) enrollment to 12 weeks and (ii) enrollment to 24 weeks: These two outcomes will be defined as the 12-week WAZ value minus the baseline WAZ value and the 24-week WAZ value minus the baseline WAZ value.

5) Mean change in LAZ from (i) enrollment to 12 weeks and (ii) enrollment to 24 weeks: These two outcomes will be defined as the 12-week LAZ value minus the baseline LAZ value and the 24-week LAZ value minus the baseline LAZ value.
6) Mean change in tricep skinfold thickness measurements from (i) enrollment to 12 weeks and (ii) enrollment to 24 weeks: These two outcomes will be defined as the 12-week skinfold thickness measurements value minus the baseline skinfold thickness measurements value and the 24-week skinfold thickness measurements value minus the baseline skinfold thickness measurements value.
7) Severe acute malnutrition (SAM) (i) enrollment to 12 weeks and (ii) enrollment to 24 weeks. SAM will be defined as a WLZ < -3 z-score or MUAC <115 mm or the presence of bilateral pitting oedema.
8) All-cause mortality (i) enrollment to 12 weeks and (ii) enrollment to 24 weeks: Mortality will include deaths from any cause. Deaths that occur on or before day 84 of follow-up will be counted in the 12-week outcome, and all deaths that occur on or before day 168 of follow-up will be counted for the 24-week outcome.
9) SAM or All-cause mortality from i) enrollment to 12 weeks and (ii) enrollment to 24 weeks. This is a composite outcome of SAM or death; if SAM occurs first, the date of SAM measurement will be used to define the time of the event. SAM will be defined as a WLZ < -3 z-score or MUAC <115 mm or the presence of bilateral pitting oedema at any follow-up visit. Deaths that occur on or before day 84 of follow-up will be counted in the 12-week outcome, and all deaths that occur on or before day 168 of follow-up will be counted for the 24-week outcome.
10) Any hospitalization from i) enrollment to 12 weeks and (ii) enrollment to 24 weeks: Hospitalization from any cause will be defined as a stay in a health facility >24 hours.

Second-Stage Randomization Intervention: Sustained Recovery Intervention vs Standard of Care

Primary Outcome

1) Sustained anthropometric recovery will be defined as meeting the 12-week anthropometric outcome by 12 weeks and continuing to have an anthropometric recovery outcome at 24 weeks. Note that all children who are enrolled in the second-stage intervention trial will have recovered by 12 weeks due to the enrollment criteria; therefore, sustained recovery is based on the 24-week measurement.

   If a child was enrolled in the trial due to WLZ criteria alone (WLZ <-2 but ≥ -3), then sustained recovery will be defined as having a WLZ ≥ -2.0 at 12 weeks to meet the 12-week anthropometric recovery outcome and a WLZ ≥ -2.0 at the 24-week visit. If a child was enrolled due to MUAC criteria alone (MUAC <125 mm but ≥115mm), then sustained recovery will be defined by having a MUAC ≥ 125 mm by 12 weeks for the 12-week anthropometric recovery outcome and a MUAC ≥ 125 mm at the 24-week visit. If a child was enrolled due to both WLZ and MUAC criteria, then sustained recovery is defined as the child having either WLZ ≥ -2.0 or MUAC ≥ 125 mm by 12 weeks for the 12-week anthropometric recovery outcome and having either WLZ ≥ -2.0 or MUAC ≥ 125 mm at the 24-week visit.

Secondary Outcomes

1) Mean change in MUAC from 12 weeks (randomization) to 24 weeks: This will be the 24-week MUAC value minus the 12-week MUAC value.
2) Mean change in WLZ from 12 weeks (randomization) to 24 weeks: This will be the 24-week WLZ value minus the 12-week WLZ value.
3) Mean change in WAZ from 12 weeks (randomization) to 24 weeks: This will be the 24-week WAZ value minus the 12-week WAZ value.
4) Mean change in LAZ from 12 weeks (randomization) to 24 weeks: This will be the 24-week LAZ value minus the 12-week LAZ value.
5) Mean change in tricep skinfold thickness measurements from 12 weeks (randomization) to 24 weeks: This will be the 24-week skinfold thickness measurements minus the 12-week measurement value.
6) Severe acute malnutrition (SAM) from 12 weeks (randomization) to 24 weeks. SAM will be defined as a WLZ < -3 z-score or MUAC <115 mm or the presence of bilateral pitting oedema.
7) All-cause mortality from 12 weeks (randomization) to 24 weeks. Deaths that occur on or before day 168 of follow-up will be counted for the 24-week outcome.
8) SAM or All-cause mortality from 12 weeks (randomization) to 24 weeks. This is a composite outcome of SAM or death; if SAM occurs first, the date of SAM measurement will be used to define the time period. SAM will be defined as a WLZ <-3 z-score or MUAC <115 mm or the presence of bilateral pitting oedema at any follow-up visit.
9) Any hospitalization from 12 weeks (randomization) to 24 weeks: Hospitalization from any cause will be defined as a stay in a health facility >24 hours.

Combination of First and Second Randomization Interventions

In a secondary paper, we will explore the combinations of the first and second randomization interventions, which result in 6 groups (3 first-stage randomization interventions and two second-stage randomization interventions). The study may not be adequately powered to look at the difference between the 6 combinations, and therefore, we will report these as exploratory outcomes.

We will compare MDSF + sustained recovery intervention (Figure 1 Groups A+B), MDSF + standard (A+C), LAF + sustained recovery intervention (D+E), LAF + standard (D+F), RUSF + sustained recovery intervention (G+H), RUSF+ standard (G+I). We will use weighting methods to take into account the SMART study design to address 'double counting' of the groups that did not recover at 12 weeks (Group A, Group D, Group G) (Nahum-Shani I, et al).

Exploratory Outcomes

1) 24-week sustained anthropometric recovery: Sustained anthropometric recovery is defined as meeting the anthropometric recovery by (required for the second stage randomization) and continuing to meet the recovery outcome at 24 weeks. Given that the second-stage randomization occurs after meeting the 12-week anthropometric recovery outcome, the analysis for 24-week recovery for the sustained recovery intervention vs standard of care will only utilize data from the 24-week visit. If a child was enrolled in the trial due to WLZ criteria alone (WLZ <-2 but ≥-3), then sustained recovery will be defined as a WLZ ≥ -2.0 at the 24-week visit. If a child was enrolled due to MUAC criteria alone (MUAC

<125 mm but ≥115mm), then sustained recovery will be defined by having a MUAC ≥ 125 mm at 24-weeks. If a child was enrolled due to both WLZ and MUAC criteria, then sustained recovery is defined as the child having either WLZ ≥ -2.0 or MUAC ≥ 125 mm at 24-weeks.

2) Mean change in MUAC from randomization to 24 weeks: This will be the 24-week MUAC value minus the randomization MUAC value.
3) Mean change in WLZ from randomization to 24 weeks: This will be the 24-week WLZ value minus the randomization WLZ value.
4) Mean change in WAZ from randomization to 24 weeks: This will be the 24-week WAZ value minus the randomization MUAC value.
5) Mean change in LAZ from randomization to 24 weeks: This will be the 24-week LAZ value minus the randomization LAZ value.
6) Mean change in skinfold thickness measurements from randomization to 24 weeks: This will be the 24-week skinfold thickness measurements minus the randomization measurement value.
7) Severe acute malnutrition (SAM) from randomization to 24 weeks. SAM will be defined as a WLZ < -3 z-score or MUAC <115 mm or the presence of bilateral pitting oedema.
8) All-cause mortality from randomization to 24 weeks (randomization) to 24 weeks. Deaths that occur on or before day 168 of follow-up will be counted for the 24-week outcome.
9) SAM or All-cause mortality from randomization to 24 weeks. This is a composite outcome of SAM or death; if SAM occurs first the date of SAM measurement will be used to define the time period. SAM will be defined as a WLZ < -3 z-score or MUAC <115 mm or the presence of bilateral pitting oedema at any follow-up visit. Deaths that occur on or before day 168 of follow-up will be counted for the 24-week outcome.
10) Any hospitalization to 24 weeks: Hospitalization from any cause will be defined as a stay in a health facility >24 hours.

3.3 Trial Sample Size

The trial sample size was based on sample size calculations for the primary outcome analyses. The sample size required for the analysis of non-inferiority of LAF as compared to RUSF on 12-week anthropometric recovery was the largest and therefore determined the overall trial sample size.

Primary Comparison #1- LAF vs RUSF on 12-week anthropometric recovery (non-inferiority): There is some evidence that justifies a non-inferiority comparison between LAF and RUSF. A meta-analysis on the management of acute malnutrition in LMICs suggests no difference in recovery from moderate wasting when standard RUSF was compared to local or homemade food, although this is considered low-quality evidence from only two studies.

The choice of non-inferiority margin was based on the judgment of clinical and public health importance of differences in the primary outcome (in our case, anthropometric recovery by 12 weeks of treatment). The investigator group and WHO staff determined that if the recovery rate with one treatment option was 60% and a non-inferiority margin of a risk difference of -5% is appropriate.

Based on assumptions of 1:1 randomization, 90% power, a one-sided α=0.025 (0.05/2 for multiple comparisons), 5% loss to follow-up/missing data, and that 60% of participants would have 12-week anthropometric recovery outcome in the RUSF and LAF arms, we would require 2,120 participants in each intervention group. As the rate-limiting randomization #1 sample size, the trial 2,120 participants are enrolled in each of the three first-stage randomization interventions, totaling 6,360 total participants in the trial.

Primary Comparison #2: MDSF vs RUSF on 12-week anthropometric recovery (superiority): This hypothesis is based on the evidence from a proof-of-principle study conducted in Bangladesh, which showed that MDSF is better than RUSF in promoting growth. We anticipate that 60% of enrolled children will achieve anthropometric recovery within 12 weeks of enrolment in the RUSF group (comparison group). Based on an enrollment of 2,120 per intervention arm, 1:1 randomization, an α of 0.025 (0.05/2 due to multiple comparisons), and 5% loss to follow-up/missing data, we will have 90% power to detect a relative risk of 1.09 for 12-week anthropometric recovery for MSDF as compared to RUSF.

Primary Comparison #3 MDSF vs RUSF on 24-week sustained anthropometric recovery (superiority): This hypothesis is also based on the evidence from a proof-of-principle study conducted in Bangladesh, which showed that MDSF is better than RUSF in promoting growth.

We anticipate that 40% of enrolled children in the second randomization, therefore 24% overall will achieve 24-week sustained anthropometric recovery in the RUSF group (comparison group). Based on an enrollment of 2,120 per arm, 1:1 randomization, an α of 0.025 (0.05 / 2 due to multiple comparisons), and 10% loss to follow-up/missing data, we will have 90% power to detect a relative risk of 1.21 for 24-week sustained anthropometric recovery for MSDF as compared to RUSF.

Primary Comparison #4 LAF vs RUSF on 24-week sustained anthropometric recovery (superiority): We believe that the LAF group could have more sustained recovery than RUSF because it has higher acceptability, lower likelihood of displacing usual diet and breastfeeding, and an easier transition to giving adequate home foods to the child in phase 2 after recovery has been achieved.

We anticipate that 40% of enrolled children in the second randomization, therefore 24% overall will achieve 24-week sustained anthropometric recovery in the RUSF group (comparison group). Based on an enrollment of 2,120 per arm, 1:1 randomization, an α of 0.025 (0.05 / 2 due to multiple comparisons), and 10% loss to follow-up/missing data, we will have 90% power to detect a relative risk of 1.21 for 24-week sustained anthropometric recovery for LAF as compared to RUSF.

Primary Comparison #5 Sustained recovery intervention vs standard of care on 24-week sustained anthropometric recovery (superiority): We believe that a follow up intervention is necessary to sustain recovery from moderate wasting, as the child has not achieved full immunological and metabolic recovery (despite anthropometric recovery) and is therefore still vulnerable and requires continued counselling and the family needs to increase intake of home foods especially where there is household food insecurity.

We assume that among participants who had 12-week anthropometric recovery (estimated 60%) and are therefore included in the second-stage randomization (n=1272 in each arm are estimated to undergo the second randomization). We assume that 45% in the standard of care group will have 24-week sustained anthropometric recovery. Based on an enrollment of 1,272 per arm, 1:1 randomization, an $\alpha = 0.05$, and 5% loss to follow-up/missing data, we will have 90% power to detect a relative risk of 1.15 for 24-week sustained anthropometric recovery for the sustained recovery intervention as compared to the standard of care.

4 General Analysis Considerations

4.1 Hypothesis Testing Framework for Primary Outcomes

The study includes multiple hypothesis tests, which are presented separately for Phase 1 and Phase 2 interventions.

First Randomization Interventions: MDSF vs LAF vs RUSF

1. We hypothesize that supplementation with MDSF will be superior to supplementation with RUSF, in the proportion of moderately wasted children showing initial anthropometric recovery by week 12 post-enrolment (ii) and superior to RUSF on the proportion showing sustained recovery by week 24 post-enrolment.

2. We hypothesize that supplementation with (i) LAF will be non-inferior to supplementation with RUSF, on the proportion of moderately wasted children showing initial anthropometric recovery by week 12 post-enrolment (ii) and superior to RUSF on the proportion showing sustained recovery by week 24 post-enrolment.

Second Randomization Interventions: Sustained recovery intervention vs Standard of Care

We hypothesize that the sustained recovery intervention will result in superior rates of sustained recovery at 24 weeks.

4.2 Primary ITT Analysis Populations

The trial will use an intent-to-treat analysis, and therefore, all randomized participants will contribute to the group they were randomly assigned. For first-stage interventions (MDSF, LAF, RUSF), the analysis population will include all participants who were randomized to a first-stage intervention. For second-stage interventions, the analysis population will include all participants who were randomized to a second-stage intervention (only individuals who recovered during Phase 1).

4.3 Missing Data

We will assume that all outcome data will be missing completely at random, and we will use a complete case analysis (only including those with data for the outcome of interest) as the primary analytic strategy. We will not impute values for any missing outcome data. We will compare the characteristics of people loss to follow-up and those with primary outcome data, we will conduct

a sensitivity analysis using inverse probability weighting (IPW) if there is indication of differential missingness/loss to follow-up by treatment arm ($p<0.05$).

4.4 Multiple Testing

We have adjusted for multiple testing for primary outcomes of the same randomization stage and outcome. For 12-week anthropometric recovery for first-stage randomization interventions, we will make two primary comparisons: MDSF vs RUSF and LAF vs RUSF. As a result, we used a Bonferroni correction and assumed an $\alpha=0.025$ (0.05/2) in sample size calculations, and we will consider a p-value <0.025 as statistically significant. For 24-week sustained anthropometric recovery for first-stage randomization interventions, we will make two primary comparisons: MDSF vs RUSF and LAF vs RUSF. As a result, we used a Bonferroni correction and assumed an $\alpha=0.025$ (0.05/2) in sample size calculations, and we will consider a p-value <0.025 as statistically significant. There is only one test of the second-stage randomization intervention, and therefore no correction for multiple testing will be applied and an $\alpha=0.05$ is used in sample size calculations, and we will consider a p-value <0.05 as statistically significant.

4.5 Treatment Adherence

We will calculate treatment compliance for Phase 1 interventions (MDSF, LAF, RUSF) based on three metrics.

- % of days trial regimen consumed: The % of days that the trial regimen (any amount) was consumed based on regimen logs and participant reports. The denominator will be 84 for those that complete follow or total follow-up days for those with withdrawal, death, censoring. Based on this metric, participants will not be able to exceed 100% adherence if they consume more of the intervention than intended.
- % of daily supplements consumed over follow-up: This will be the average % of the supplement consumed each day of follow-up (e.g. 80% Day 1, 100% Day 2, 0% Day 3 etc.). The number of averaged will be 84 for those that complete follow or total follow-up days for those with withdrawal, death, censoring.
- % of daily supplements consumed among days the child received any supplement: This will be the average % of the supplements consumed on days they receive the regimen, (e.g. 80% Day 1, 100% Day 2, etc.).

4.6 Timing of Analyses and DSMB Stopping Guidance

Interim analyses for the primary outcomes and severe adverse events will be presented to the DSMB by the study arm (but coded to not reveal specific intervention – i.e., red, blue, green groups) per the DSMB charter. For the primary outcomes, the Haybittle-Peto boundary ($p<0.001$) will be used by the DSMB to guide stopping recommendations. Stopping the trial will also be determined by the DSMB but also considered by the DSMB based on severe adverse event rates, recruitment, and follow-up rates, new external information, or other considerations.

The final analysis will be conducted after all participants have been discharged from the study.

4.7 Pilot Data

Each study site will conduct a pilot of 12 participants before starting the trial. The participants in the pilot will not be analyzed in the trial analysis.

5 Statistical Analysis

5.1 Screening Data

We will report the trial start and end dates for screening as well as the overall number of children screened, the number of screened children and not recruited the reason for non-recruitment, and the number of children enrolled in the trial in a standard trial flow chart. We will also produce a table of these data by study site, which may be considered for a supplemental table in the main trial report.

5.2 Statistical Methods for Primary Outcomes

i. **MDSF vs RUSF on 12-week anthropometric recovery (superiority):** We will use data from all follow-ups visits up to 12-week visit to define a 12-week anthropometric recovery for each participant. Participants who miss study visits, die, or withdraw from the study will have data from all completed study visits analyzed. If a participant does not complete any follow-up visits at or before the 12-week visit (no follow-up data), they will be excluded from the analysis (complete case). All individuals who have at least 1 follow-up visit will contribute to the analysis. We will not impute missing values for missed visits; the definition of recovery will be based on data available from visits at or before 12 weeks.

   We will report the proportion of participants who reached 12-week anthropometric recovery by intervention arm. We will use a log-binomial regression model to present the relative risk (RR) of 12-week anthropometric recovery for the MDSF arm as compared to the RUSF arm. Due to stratified randomization by study site, we will include a fixed effect for the study site in the model. We will present the relative risk, two-sided 97.5% confidence intervals, and a p-value <0.025 will be considered statistically significant (to account for multiple comparisons).

ii. **LAF vs RUSF on 12-week anthropometric recovery (non-inferiority):** This comparison assesses first-stage randomization options. We will use data from all completed follow-up visits before and at the 12-week visit to define a 12-week anthropometric recovery for each participant. Participants who miss study visits, die, or withdraw from the study will have data from all completed study visits analyzed. If a participant does not complete any follow-up visits at or before the 12-week visit (no follow-up data), they will be excluded from the analysis (complete case). We will not impute missing values for missed visits; the definition of recovery will be based on data available from visits at or before 12 weeks.

We will report the proportion of participants who reached 12-week anthropometric recovery by LAF and RUSF arms. We will use generalized linear models with a binomial distribution and identity link function to produce risk difference estimates for the LAF arm compared to the RUSF arm. Due to stratified randomization by study site, we will include a fixed effect for the study site in the model. The lower bound of the one-sided 97.5% confidence interval (to account for multiple comparisons) of the risk difference arm will be compared with the prespecified noninferiority margin of -5% to assess non-inferiority. We will also present a p-value for non-inferiority, with a p-value <0.025 considered statistically significant. Two-sided 95% confidence intervals will be presented, which have an upper boundary equivalent to a one-sided 97.5% confidence interval.

iii. MDSF vs RUSF on 24-week sustained anthropometric recovery (superiority): This analysis will utilize data from visits during the first 12 weeks in addition to the 24-week visit. Sustained anthropometric recovery is defined by 12-week anthropometric recovery and continued recovery at the 24-week visit.

If a participant does not complete any follow-up visits at or before the 12-week visit (no 12-week follow-up data), they will be excluded from the analysis (complete case). Individuals who reached anthropometric recovery at 12 weeks but do not have anthropometric data at 24 weeks (loss to follow-up/missing data, death, withdrawal) will also be excluded from the analysis (complete case). Individuals who did not reach anthropometric recovery by 12 weeks will be analyzed as not having met 24-week sustained recovery, regardless of completion of the 24-week visit, as by definition, they will not have met sustained recovery.

We will report the proportion of participants who reached sustained anthropometric recovery at 24 weeks by intervention arm. We will use a log-binomial regression model to present the relative risk (RR) of 24-week sustained anthropometric recovery for the MDSF arm compared to the RUSF arm. Due to stratified randomization by study site, we will include a fixed effect for the study site in the model. We will present the relative risk, 97.5% confidence intervals, and a p-value <0.025 will be considered statistically significant (to account for multiple comparisons).

iv. LAF vs RUSF on 24-week sustained anthropometric recovery (superiority): This analysis will utilize data from visits during the first 12 weeks in addition to the 24-week visit. Sustained anthropometric recovery is defined by 12-week anthropometric recovery and continued recovery at the 24-week visit.

If a participant does not complete any follow-up visits at or before the 12-week visit (no 12-week follow-up data), they will be excluded from the analysis (complete case). Participants who reached anthropometric recovery at 12 weeks but do not have anthropometric data at 24 weeks (loss to follow-up/missing data, death, withdrawal) will also be excluded from the analysis (complete case). Participants who did not meet 12-week anthropometric recovery will be analyzed as not having met 24-week sustained

recovery, regardless of their completion of the 24-week visit, as by definition they are known to not have met the sustained recovery endpoint.

We will report the proportion of participants who reached sustained anthropometric recovery at 24 weeks by intervention arm. We will use a log-binomial regression model to present the relative risk (RR) of 24-week sustained anthropometric recovery for the LAF arm compared to the RUSF arm. Due to stratified randomization by study site, we will include a fixed effect for the study site in the model. We will present the relative risk, 97.5% confidence intervals, and a p-value <0.025 will be considered statistically significant (to account for multiple comparisons)

v. Sustained recovery intervention vs standard of care on 24-week sustained anthropometric recovery (superiority):

Given that the second randomization occurs after meeting the 12-week anthropometric outcome, the analysis for 24-week sustained recovery for sustained recovery intervention vs standard of care will only utilize data from the 24-week visit to define the outcome for this comparison.

We will report the proportion of participants who reached sustained anthropometric recovery at 24 weeks by intervention arm. We will use a log-binomial regression model to present the relative risk (RR) of 24-week sustained anthropometric recovery for the sustained recovery intervention as compared to the standard of care arms. Due to stratified randomization by study site, we will include a fixed effect for the study site in the model. We will present the relative risk, 95% confidence intervals, and a p-value <0.05 will be considered statistically significant.

5.3 Statistical Methods for Secondary Outcomes

We present the methods for the secondary analyses by analysis type.

i. *Time-to-event analyses for first-stage and second-stage intervention effects*

The following secondary outcome(s) will be analyzed with time-to-event analyses, such as time to anthropometric recovery within the first 12 weeks.

We will use Cox proportional hazard regression models that include a fixed effect for study site to account for stratified randomization to produce hazard ratios and 95% confidence intervals. The time metric for time to event analyses will be follow-up days.

ii. *Effect of first-stage and second-stage intervention effects on the change of anthropometric outcomes*

We will use the analysis of covariance (ANCOVA) modelling approach, which includes the baseline score as a covariate. ANCOVA is known to have greater statistical power as compared to analysis of change from baseline [outcome is follow-up minus baseline] (Van

Breukelen 2006). Further, in an ANCOVA analysis, either the follow-up score or the change score can be used as the outcome variable, but using the change score will provide a more precise estimate if the correlation between baseline and follow-up is high (Fu and Holmer 2015). In our trial, we expect reasonably high correlation between baseline and follow-up anthropometric outcomes and will therefore use the change in anthropometric outcomes in the analysis and include baseline score as a covariate.

We will use generalized linear models with a change in anthropometric measures as the dependent variable. We will include intervention, baseline value (ANCOVA approach), and a fixed effect for study site as independent variables. We will present mean differences and 95% confidence intervals for the mean difference in change in anthropometric measures between intervention arms.

iii. *Effect of first-stage and second-stage intervention effects on non-repeatable binomial* We will use log-binomial regression models to present the relative risks (RR) of non-repeated secondary outcomes between groups. These outcomes include mortality, hospitalization, etc. Due to stratified randomization by study site, we will include a fixed effect for study site in the model. We will present the relative risk estimates and 95% confidence intervals.

5.4 Sensitivity Analyses for Potential Baseline Imbalance

Standard Table 1 of participant characteristics by treatment arm will not present p-values for differences between groups. In sensitivity analyses for the primary outcomes to assess the potential for baseline imbalance to affect results, we will adjust for baseline factors that showed some degree of imbalance between treatment groups based on a $p<0.10$. We will report the results of these sensitivity analyses and discuss if there are any qualitative differences in the findings of the baseline-adjusted and the primary analyses (unadjusted for baseline factors). If there is no evidence of baseline imbalance for the primary outcomes, we will not present baseline-adjusted analyses for secondary outcomes.

5.5 Per Protocol Analyses

We will conduct per-protocol analyses for the effect of first-stage randomization groups on primary and secondary outcomes. A per-protocol analysis is of particular importance for the LAF versus RUSF on the 12-week anthropometric recovery outcome, as suboptimal adherence may bias estimates to the null, which is anti-conservative in the non-inferiority design. Nevertheless, for consistency in analyses of first-stage interventions, we will present per-protocol analyses for all primary and secondary outcomes for the MDSF, LAF, and RUSF comparisons. The per-protocol analysis will include participants who take their randomized regimen for at least 75% of days between randomization and 12 weeks and have data (non-missing) for the outcome of interest at 12 or 24 weeks.

Per protocol analyses are not planned for the second-stage interventions due to the superiority testing focus and the nature of the sustained recovery intervention and control interventions.

5.6 Exploratory Analyses of Potential Effect Modifiers

We will explore the effect modification of the ITT treatment effects on the *primary outcomes* for Phase 1 and Phase 2 interventions by:

1) Site region (South Asia [Bangladesh, India, Pakistan] vs sub-Saharan Africa [Tanzania and Mali]
2) Infant age at enrollment (6 to <12.0 months vs 12.0 to 24 months)
3) Infant sex (male versus female)
4) MAM diagnosis type (WLZ only, MUAC only, both WLZ and MUAC).
5) Diarhrea at baseline (Diarrhea vs no diarrhea)

Due to the exploratory nature of these analyses, we will not adjust for multiple comparisons but rather recognize there is a risk of type II error. In terms of statistical analyses, the same general modeling approaches will be used for the primary outcomes but will include interaction terms between treatment and the effect modifier of interest. The statistical significance of effect modification will be assessed with the likelihood ratio test.

5.7 Harms

All severe adverse events (SAEs) by trial arm will be reported in the primary manuscript by treatment arm. We will also report anaphylaxis reactions by intervention arm.

5.8 Statistical Software

All analyses will be conducted with STATA and SAS.

References

Almirall, D., I. Nahum-Shani, N. E. Sherwood, and S. A. Murphy. 2014. 'Introduction to SMART designs for the development of adaptive interventions: with application to weight loss research', *Transl Behav Med*, 4: 260-74.

Fu, R., and H. K. Holmer. 2015. 'AHRQ Methods for Effective Health Care.' in, *Change Score or Followup Score? An Empirical Evaluation of the Impact of Choice of Mean Difference Estimates* (Agency for Healthcare Research and Quality (US): Rockville (MD)).

Van Breukelen, G. J. 2006. 'ANCOVA versus change from baseline: more power in randomized studies, more bias in nonrandomized studies ', *J Clin Epidemiol*, 59: 920-5.

Nahum-Shani I, Qian M, Almirall D, Pelham WE, Gnagy B, Fabiano GA, Waxmonsky JG, Yu J, Murphy SA. Experimental design and primary data analysis methods for comparing adaptive interventions. Psychol Methods. 2012 Dec;17(4):457-477.