The effects of single and combined water, sanitation, handwashing and nutrition interventions on child development in young Kenyan children

Plan for statistical analysis

Patricia Kariger, Christine P. Stewart, Lia Fernald, Amy Pickering, Charles D. Arnold, Benjamin F. Arnold, Holly N. Dentz, Audrie Lin, Theodora J Meerkerk, Erin Milner, Jenna Swarthout, John M Colford, A. Clair Null

Date Created: 16 December 2016

Funding Source: The Bill and Melinda Gates Foundation

Study Registration: ClinicalTrials.gov Identifier: NCT01704105

Study Protocol:
Arnold BF, Null C, Luby SP, Unicomb L, Stewart CP, Dewey KG, et al. 2013. Cluster-randomised controlled trials of individual and combined water, sanitation, hygiene and nutritional interventions in rural Bangladesh and Kenya: the WASH Benefits study design and rationale. BMJ Open 3: e003476
Amendment available: https://osf.io/7urqa/

1. Study Background and Rationale

Growth and development can be compromised when children lack access to good sources of nutrition and safe and clean water, sanitation and hygiene (WSH) conditions [1]. Growth faltering and micronutrient deficiencies have been associated with worse cognitive functioning, which can persist even after nutrition interventions have been provided [1, 2]. Additionally, there is some evidence that multiple bouts of diarrhea early in life are associated with both growth faltering and, subsequently, impaired cognitive development [3].

It is proposed that the failure for nutritional treatments to correct linear growth deficits and improve other child outcomes may be due to nutrient loss caused by repeated infections for children living in poor WSH environments. It is further hypothesized that these repeated infections, even if subclinical, can change gut functioning, resulting in increased compromise in nutrient absorption, growth and development [4, 5].

A recent review on the effects of WSH interventions on growth found only small gains in length in children under age five years [6]. One study has demonstrated that handwashing interventions can improve cognitive, motor and personal-social outcomes five years post-exposure [7]. Further exploration of the effects of singular and combined WSH and nutrition interventions on child growth and development are needed.

2. Study objectives and hypotheses

The WASH Benefits study measures the impact of individual and combined water, sanitation, and hygiene and nutrition interventions on growth, development and enteric disease in young children in Kenya. This document outlines the analysis plan for the child development outcomes. We aim to test the following hypotheses:

H1 (primary hypothesis): Interventions improving water quality (W), sanitation (S), hygiene (H), WSH in combination, nutrition (N), or WSHN in combination will improve indicators of child development, including measures of communication, gross motor, personal social skills, WHO motor milestones, and child head circumference.

H2: The combination of WSHN will improve child development measurements more than WSH, or N alone.

3. Study design

The WASH Benefits trial has a cluster-randomized design. In Kenya, clusters were defined as one or two, and in rare cases three, adjoining villages with at least six eligible pregnant women. Nine clusters were used to form a geographic block for the purposes of randomization. The trial had six intervention arms, a double-sized active control arm, and a single-sized passive control arm. The active control arm included visits by a health promoter and child arm circumference measurements while the passive control arm was not visited or measured by a health promoter.

4. Participants

The WASH Benefits Kenya trial was conducted in the Kakamega, Bungoma, and Vihiga counties of Western Kenya. The trial enrolled participants according to the following inclusion and exclusion criteria.

Inclusion criteria:

- At the cluster level:
    - Located in rural area
- At the individual/household level:
    - Pregnant mother in her second or third trimester (Kenya)

Exclusion criteria:

- At the cluster level,
    - High access to piped water or had chlorine dispensers in the community (Kenya)
    - No active WASH or Nutrition programs, other than those operating nationally

- At the individual/ household level

- Plans to move in the following year (excluded in Kenya to minimize loss to follow-up)
- Household does not own their home (excluded in Kenya to minimize loss to follow-up)

The children born to the enrolled pregnant mothers were considered "target" children and are the focus of this analysis.

5. Intervention arms

The WASH Benefits intervention arms included water treatment, sanitation, handwashing, nutrition, combined water treatment, sanitation and handwashing (WSH) and combined WSH plus nutrition.

- **Active Control (AC):** Monthly visits by a health promoter to measure mid-upper arm circumference (MUAC) and referrals for severe acute malnutrition (SAM). (These activities were a part of all groups below, except for the passive control).
- **Passive Control (PC):** No health promoters or intervention activities
- **Water treatment (W)**: Chlorine dispensers installed at community water locations plus bottled chlorine provided to enrolled compounds. Behavior change activities focused on treating drinking water for children aged <36 months.
- **Sanitation (S):** Pit latrine construction or upgrades with plastic slab and drop hole cover, child potties, sani-scoop to remove feces from households and compounds. Behavior change activities focused on the use of latrines for defecation, the removal of human and animal feces from the compound, safe disposal of feces.
- **Handwashing (H):** Dual dispenser tippy-tap devices installed near the latrine and kitchen area. Each provided one container for soapy water and one container for rinse water and were operated with independent pedals. Soap was provided at regular intervals. Behavior change activities focused on handwashing with soap at critical times around food preparation, after use of toilet, after cleaning the child's anus, and other contact with feces
- **Water, Sanitation, Hygiene (WSH)** combined all of the above activities.
- **Nutrition (N):** Provision of lipid-based nutrient supplements (LNS) for infants aged 6-24 months. Behavior change activities focused on age-appropriate recommendations on maternal nutrition and infant feeding practices, including dietary diversity during pregnancy and lactation, early initiation of breastfeeding, promotion of exclusive breastfeeding, timely introduction of complementary foods, dietary diversity during complementary feeding, feeding frequency, and feeding during illness.
- **WSH+Nutrition (WSHN):** All of the above

Local promoters were nominated by communities and received extensive training on arm-specific intervention activities. They visited study compounds at least monthly to deliver behavior change interventions, provide support to households in the use of the study hardware, and/or provide monthly supplies of LNS.

6. Outcomes

This study utilizes measures of child development, including communication, gross motor and personal social skills, WHO motor milestones, and child head circumference, which were gathered as secondary outcomes. Details on the statistical analysis plan for the main study outcomes have been published [8, 9]. The study analysis plan for the child development outcomes will be generally similar to the main study outcomes (child growth and diarrhea) and are therefore not described in detail here. This document will present detailed information on the outcome variables and describe where there are differences in the analytic approach from the previously published plans.

The child communication, gross motor, personal-social and global development scores will be derived from the Extended Ages and Stages Questionnaires (EASQ; adapted from Squires and Bricker, 1997) [10], a parent report measure of child developmental progression. All items were read directly to the respondents, and children were asked to demonstrate some behaviors that may be difficult to observe by the parent. The specific outcomes will be scores for each of the four scales at the Year 2 follow-up when children were 18-30 months of age.

The measurements of motor milestone achievement were collected at the Year 1 follow-up visit when children were 6-18 months of age, using a questionnaire designed to capture the developmental milestones recommended by WHO [11]. Lastly, child head circumference was measured by trained anthropometrists as part of the anthropometric assessment at the Year 1 follow-up and the Year 2 follow-up.

6.1 Definition of outcome variables for this analysis

6.1.1 EASQ

Four different EASQ age-specific questionnaires were used. Form G is for children 18-19 months of age; Form H is for children 20-21 months of age; Form I is for children 22-24 months of age; and Form J is for children 25-30 months of age. For all items asked, respondents could reply with one of three responses: *Yes, Sometimes* or *Not Yet*. *Yes* responses were assigned 10 points, *Sometimes* responses 5 points, and *Not Yet* responses 0 points.

Scores will be summed to determine a total raw score for each scale (communication, gross motor and personal-social skills) within each age group/form. A total EASQ raw score (combining all three scales) within each age group/form will also be computed. While there is some overlap in the content of items administered within each form, they each differ slightly, and thus have different numbers of items:

- **Communication scale score:**
    - Form G raw score includes items C16.C23-C16.C29 (raw scores range 0-70);

    - Form H raw score includes items C16.C25-C16.C32 (raw scores range 0-80);
    - Form I raw score includes items C16.C27-C16.C34 (raw scores range 0-80);
    - Form J raw score includes items C16.C27-C16.C36 (raw scores range 0-100).
- **Gross motor scale score:**
    - Form G raw score includes items C16.M24-C16.M31 (raw scores range 0-80);
    - Form H raw score includes items C16.M26-C16.M32 (raw scores range 0-70);
    - Form I raw score includes items C16.M28-C16.M35 (raw scores range 0-80);
    - Form J raw score includes items C16.M28-C16.M36 (raw scores range 0-90).
- **Personal-social scale score**
    - Form G raw score includes items C16.P23-C16.P30 (raw scores range 0-60);
    - Form H raw score includes items C16.P25-C16.P33 (raw scores range 0-70);
    - Form I raw score includes items C16.P27-C16.P34 (raw scores range 0-60);
    - Form J raw score includes items C16.P27-C16.P38 (raw scores range 0-100).
- **Global EASQ scale score, (combined score from all three scales)**
    - **Form G global total raw score includes** items C16.C23-C16.C29, C16.M24-C16.M31**,** C16.P23-C16.P30 (raw scores range 0-210)
    - **Form H global total raw score includes** C16.C25-C16.C32, C16.M26-C16.M32, C16.P25-C16.P33 (raw scores range 0-220)
    - **Form I global total raw score includes** C16.C27-C16.C34, C16.M28-C16.M35, C16.P27-C16.P34 (raw scores range 0-220)
    - **Form J global total raw score includes** C16.C27-C16.C36, C16.M28-C16.M36, C16.P27-C16.P38 (raw scores range 0-290)

To create the reference distributions for each scale (communication, gross motor, personal-social and global score), the raw scores from the double-sized Active Control group alone will be standardized with a mean of 0 and a standard deviation of 1 yielding Z-scores for each age band within the Active Control group. Standardized Z-scores for the other (treatment and passive control) groups will then be created using the reference distribution for each age band/group. Reference distributions for each scale will be created by form using the age bands corresponding with each form. Since form J has a larger age range than other forms, it will be divided into two age bands. There is also a subset of children for whom the wrong form was completed (form I was completed, but based on the children's age form J should have been done). Since there are overlapping questions between forms I and J, data from these children can be used, but they will be in a separate category. Therefore, the final categories for creating the reference distributions for each scale are as follows:

Category	Form	Age range (months)
1	G	17.6 - 19.5
2	H	19.6 – 21.5
3	I	21.6 – 24.5
4	J	24.6 – 27.5
5	J	>27.6
6	I, should have completed J	24.6-27.5

In the Active Control group, there is a small sample size (<30) in the groups of children completing form G and completing form I who should have completed form J. Therefore, all children in these two categories will be removed from analyses (n=91).

6.1.2 Motor Milestones

The motor milestone questions were adapted from the WHO Motor Development Study [11] and asked during the Year 1 follow-up survey (C16.1 – C16.6) for study children. The questions asked about six behaviors: 1) sitting without support; 2) hands-and-knees crawling; 3) standing with assistance; 4) walking with assistance; 5) standing alone; and 6) walking alone. Caregivers were shown pictures of each milestone, and were asked to report if the child is able to do the behavior and if so, whether they were observed doing the behavior in the last 24 hours. The primary indicator for this analysis will be the caregiver report of each behavior ever being achieved (column A from the questionnaire).

6.1.3 Head Circumference Z-scores

Child head circumference was measured in triplicate at the Year 1 and Year 2 follow-up survey by trained and standardized anthropometrists using non-stretchable 65 cm insertion tapes (Weigh and Measure LLC, Olney, MD). Anthropometrists positioned the tape just above the supraorbital ridge in the front and the occiput at the back of the skull. Parents were asked to remove any hair ornaments from the child prior to measurement. The tape was tightened to compress the hair and read from the side. Measurements were performed in triplicate. The median value will be used for the analysis. Head circumference for age z-scores will be calculated using the WHO growth standards [12]. All anthropometrists received training and were standardized before each survey. Replicate measurements were collected during the course of data collection on a subsample of children measured by each anthropometrist to check for error and bias in the measurements.

6.2 Primary and secondary outcomes

Our primary outcome measures for this analysis will include the global mean EASQ z-score, head circumference, and age of motor milestone acquisition.

Our secondary outcomes will include the mean EASQ subscales and the prevalence of low head circumference (<-2 z-scores).

7. Minimum detectable effect size

The rationale for the sample size for the main trial is described in detail in Arnold et al, 2103 [8]. In brief, the sample size was chosen to detect a difference of 0.15 in LAZ, assuming a type I error (α) of 0.05, power ($1-\beta$) of 0.8, and a 10% dropout after baseline. The control arm was chosen to be double sized to account for the multiple hypothesis tests. We would expect a similar detectable difference for the EASQ z-scores in this analysis.

8. Effect modifiers

- **Child sex:** Biological differences, differential care practices, or other behavioral practices may modify the effect of the interventions on girls vs boys.
- **Child birth order:** First born (mother primiparous) vs. second or greater (mother multiparous)
- **Maternal age**: Biological differences, differential care practices, or differential social support networks may modify the effect of the interventions. We will utilize a cutoff of ≤20 years to define young maternal age.
- **Maternal education**: Maternal education has been strongly associated with child development outcomes and child care practices in many studies. It is possible that the intervention effect may vary by maternal education level. We will utilize a cutoff of completion of primary schooling to define "low education" and secondary and above as "high education"
- **Food insecurity:** Children living within households with a greater degree of food insecurity are more vulnerable to micronutrient malnutrition and anemia. We would hypothesize that they would also have the potential to benefit most from the nutritional supplementation intervention. In Kenya, we have used the 3-item Household Hunger Scale (HHS) as our indicator of food insecurity [13]. The HHS is typically grouped into three categories: 1) little to no hunger, 2) moderate hunger, or 3) severe hunger. We will dichotomize this score into little to no hunger vs moderate to severe hunger.
- **Socioeconomic status:** Using the socioeconomic score, we will categorize households into quintiles and then dichotomize into the highest quintile vs. the lowest four quintiles.

9. Analysis principles

<u>9.1 Ceiling and floor effects</u>

Ceiling (i.e., passing all items) or floor (i.e., not passing any items) effects for the EASQ scales are possible. This would result in a truncated distribution at the upper and lower tails and lumping of values at the highest and lowest possible scores. While our pilot and preliminary analyses suggest a low rate of children scoring at the floor of the test, there may be 10%-15% of children at the ceiling of some of the scales. To handle this, we will consider a few approaches:

- If the prevalence of scores at the ceiling in the control group is observed to be low (i.e. ≤10%) across all three scales and the global scale within any age group/form, we will use mean score comparisons of treatment groups to control group.
- If the prevalence of scores at the ceiling in the control group is observed to be >10% for any one of the three scales or the global scale within an age

group/form, we will exclude that subscale for that age group from the analysis, but analyze all other data.

- If the prevalence of scores at the ceiling in the control group is observed to be >10% across all three subscales and the global scale we will consider an alternate analysis plan described in more detail below.

9.2 Hypothesis testing

To test our hypotheses, we will conduct two sided hypothesis test comparisons between the following study arms.

H1. (primary hypothesis): PC, W, S, H, WSH, N, WSHN compared to the AC group (7 comparisons)
H2. The WSHN group compared to each of the WSH, and N groups (2 comparisons)

9.2.1 EASQ and head circumference models

We will estimate unadjusted and adjusted intention-to-treat effects between study arms. Our parameters of interest for our primary outcomes will be the mean difference in head circumference, the global EASQ, communication, gross motor, and personal social-skills between the intervention groups and the active control (for H1) or between the WSHN group and the WSH or N groups (for H2). If the outcome variables follow a skewed distribution, we will explore appropriate transformations prior to analysis, such as a log transformation. We will estimate the mean difference parameter using a generalized linear model (GLM) with robust standard errors at the study block level. We will analyze data matched on randomization block.

We will check for the normality of the sample distribution using the Kolmogorov-Smirnov testThe null hypothesis will be that the sample medians are equal between the intervention groups and the active control (for H1) or between the WSHN group and the single intervention groups (for H2). The null hypothesis will be rejected if the groups are not equal.

If the proportion of children scoring at the ceiling is high as described above, we will analyze the standardized EASQ scores stratified by age category. We will also aggregate over age and analyze the data in a combined model that includes the intervention group term, age category, and an age*group term. If the average mean differences over all ages are comparable to the age-stratified estimates (e.g. within approximately 10% of each other), only the combined averages will be presented. As a sensitivity test, we will further analyze the data in a parametric right-censored regression model assuming either a normal or log-normal distribution, depending on the apparent best fit of the distribution. This approach assumes a theoretical distribution of the EASQ scores (e.g., Gaussian) among children scoring at the ceiling that may be false, so this model will not be the one from which we draw our primary inferences. Rather, it might provide a useful view of possible

biased effect size estimates in the non-parametric models due to a differential proportion of children scoring at the ceiling by group.

Head circumference will also be analyzed dichotomously using a cutoff of < -2 z-scores. Our parameter of interest for dichotomous outcomes will be the prevalence ratio and difference. We will estimate these parameters using GLM as for the continuous outcomes, using a binomial distribution and an identity link for the prevalence difference and a log link for the prevalence ratio.

9.2.2 WHO Motor Milestone models

The WHO motor milestone measurements are in the form of "current status" data, an extreme form of left-censored measurement, in which the outcome, if it has occurred, is known to have occurred by the age of measurement; yet, the specific age at which it occurred is unknown. This is a common data format in survival analysis, and well-developed methods exist for the analysis of such data [14]. A common motivating application has been the estimation of the age of weaning from cross-sectional data [15], which is a direct analog to WHO motor milestone achievement. In such cases, recall of specific ages of achievement is likely subject to error or bias, so the current status at the time of measurement is used instead. The approach is well-suited to studies where the monitoring age varies across children, as it does in the WASH Benefits trial. Let T denote the age of the child when she achieved the milestone, and let C denote the age at which she was measured. The observed outcome data is $O = (Y, C, A, W)$, where $Y = I(T \leq C)$, A is the randomized treatment assignment (equal to 1 if intervention, 0 if control), and W are baseline (pre-randomization) covariates. We assume that a child's monitoring age in days (C) is independent of the age of milestone achievement (T), which is very reasonable in this context since field logistics and intervention timing dictated monitoring times. Given the observed data and this assumption, we can estimate the cumulative distribution of T as the conditional probability of Y given C [14].

We will estimate the cumulative distribution of milestone achievement – analogous to a survival curve – for each of the six WHO motor milestones. We will fit a separate survival curve for each arm in the trial using the nonparametric maximum likelihood estimator (NPMLE) fit using the pool-adjacent-violators-algorithm (PAVA) [16] or if necessary the expectation-maximization (EM) algorithm. The NPMLE method for current status data is analogous to a Kaplan-Meier estimator of the survival curve in right-censored data in longitudinal studies. In addition to the NPMLE survival curves, we will model the rates of WHO milestone acquisition using a semi-parametric Cox-proportional hazards model, which can be estimated in current status data using a generalized linear model with complementary log-log link [14]:

$$\log - \log[1 - \mathrm{E}(Y \mid C, A, W)] = \log - \log[S_0(C)] + \beta A + \gamma W.$$

We will model the baseline hazard ($\log - \log[S_0(C)]$) with a spline in a generalized additive model [14]. The primary analysis will be unadjusted (i.e., no W). Adjusted analyses will include pre-specified covariates if they are associated with the outcome (next section).

We will compare WHO milestone achievement curves between each intervention arm and the control arm using a hazard ratio, which is estimated by exponentiating the coefficient β in the above model. Methods are implemented in the `isotone` and `mgcv` packages in R [17]. Similar to the primary analysis of the trial, we will compare each of the intervention arms against the control arm (6 tests), and will compare the combined WASH+Nutrition intervention against the combined WASH (WSH+N vs. WSH) and against the nutrition (WSH+N vs. N) arm (2 additional tests).

Finally, we will only conduct the analyses of WHO motor milestones for which there is a reasonable amount of variation in outcome within the study populations at the age of monitoring. If, for example, nearly all children (e.g. ≥95%) can sit alone without support by the age they are monitored in the year 1 follow-up survey, then we will not be able to compare groups for differences in that milestone.

9.2.3 Adjusted analysis

For each outcome above, we will conduct three sets of adjusted analyses using the covariates described below. We will pre-screen covariates to assess whether they are associated with each outcome prior to including them in adjusted statistical models. We will use the likelihood ratio test to assess the association between each outcome and each covariate and will include covariates with a p-value < 0.1 in the adjusted analysis. We will also exclude covariates that have little variation in the study population (e.g., prevalence <5%).

1) Child age in days only

2) The above plus covariates described below:

Child sex
Child birth order (defined as first born or ≥second born)
Maternal age (years)
Maternal height (cm)
Maternal education (no education or incomplete primary, complete primary, incomplete secondary, or completed secondary)
Number of children <18 years of age in household
Total number of people in compound
Food insecurity of household measured using the Household Hunger Scale
Housing materials (floor, walls, roof), categorized dichotomously as poor quality (mud, including cane/palm/bamboo with mud, or mud with stones) and higher quality (including concrete, cement, tile, bricks)
Household assets: electricity, radio, television, mobile phone, clock, bicycle, motorcycle, stove, gas cooker, car.
Animal ownership: Number of cows, goats, dogs, chickens/poultry
Distance to water source
Field staff team member who administered EASQ
Month of measurement

3) The above plus covariates listed below, with subsequent details on the variable creation. These variables have well established associations with child development scores and are often included in multivariable adjusted models in the literature [18, 19]. However, these factors were measured after randomization. While we do not expect them to be influenced by the randomized exposures, it is possible that they could have been indirectly affected by intervention components, such as health promoter visits. We will therefore consider them in separate adjusted models.

Indicators of Family Care (summed separately)
Maternal/Primary caregiver mental health (continuous scores, summed)

Home stimulation and care was assessed using the Family Care Index using a module similar to that used in the UNICEF Multi-Indicator Cluster Surveys (MICS) [20]. These items assess the availability of books and play materials for children, as well as the occurrence of various activities between caregivers and the child, such as reading, singing songs, or playing. The questionnaire module was administered on the same day as the anthropometric measurements during the 2 year visit survey round. The number of books in the home was recorded as a continuous variable and recoded to 1=any and 0=none based on the distribution of responses. The variety of play materials was assessed by asking about play with 3 items (homemade toys, manufactured toys, or other household objects). Two questions assessed the number of days that the child was left alone or in the care of another child for >1hr and responses were recorded continuously. Both variables were recoded as binary variable where 0=1-7 days and 1=0 days. Children's interactions with caregivers were assessed by asking about 6 activities (reading books, telling stories, singing songs, playing with the child, taking the child outside the home, and naming/counting/drawing with the child) by three categories of caregivers (mother, father, and other). Each of these activities was given a score of 0-3 depending on how many caregivers performed the activity with the child. A family care index will be created by summing up all of the dichotomous indicators with a range of potential scores from 0-24. Children with a missing value on any of the questions will be coded as missing on the index.

Maternal mental health was assessed using the Patient Health Questionnaire [21]. Eight questions were asked about the frequency of feelings or behaviors over a two week period. These questions will be scored such that a reported frequency of 0-1 days = 0; 2-6 days = 1; 7-11 days =2; 12-14 days = 3. Respondents who report that they are unwilling or not comfortable to respond will be counted as missing. A continuous score will be calculated by summing the responses over the 8 items. Thus, the score can range from 0-24.

9.3 Effect modification and subgroup analyses

We will examine the potential for effect modification on the additive scale by testing an interaction term between intervention group and the factors specified in section 8 above in linear models. Results will be presented in stratified tables. Effect modification will be examined by reviewing the stratified point estimates and confidence intervals to look for significant (stratified $p<0.05$) and consistent trends.

9.4 Missing data

We will tabulate the percentage of participants lost to follow-up between enrollment and the assessment of the outcomes at the one and two-year follow-up. Children who have migrated out of the study site, but who were met at either follow-up visit will be included in the analysis. We will assess whether missing data are differential with respect to treatment group or any characteristics of participants, by comparing rates of missingness across randomized arms and comparing baseline covariates between those with an available vs. missing data [22]. We will also compare the balance of baseline covariates between study arms both in the original enrolled population and among participants that successfully provided data. We will conduct a complete-case analysis, and if we find evidence of high levels of missingness (>20%; differential or non-differential by study arm), we will also conduct an inverse probability of censoring-weighted analysis that reconstructs the original enrolled population [23]. We will exclude pregnancy losses from this analysis, but will include postnatal infant and child deaths.

9.5 Blinding during analysis

We plan to conduct all analyses blinded to intervention groups by using an independently created set of randomization codes that does not match actual group assignments. Blinding will be maintained until analyses have been replicated and compared by two analysts.

9.6 Primary models

The above text has described a number of alternative approaches and models to adjust for covariates, account for attrition and missing data, and consider as sensitivity checks. We hope that the results with these adjustments will be fairly consistent across method, but we expect that point estimates and confidence intervals will vary to some degree. We will consider our primary estimates of effect to be those calculated in the unadjusted models.

10. Update history

Jan 27 2017: We removed the statistical criteria for determining effect modification due to the large number of statistical tests. Previous versions of this plan specified that we would use a p-value of 0.1 as a threshold for statistical significance.

References

1. Walker SP, Wachs TD, Gardner JM, Lozoff B, Wasserman GA, Pollitt E, Carter JA: **Child development: risk factors for adverse outcomes in developing countries.** *Lancet* 2007, **369:**145-157.
2. Grantham-McGregor S, Cheung YB, Cueto S, Glewwe P, Richter L, Strupp B: **Developmental potential in the first 5 years for children in developing countries.** *Lancet* 2007, **369:**60-70.
3. Scharf RJ, Deboer MD, Guerrant RL: **Recent advances in understanding the long-term sequelae of childhood infectious diarrhea.** *Curr Infect Dis Rep* 2014, **16:**408.
4. Humphrey JH: **Child undernutrition, tropical enteropathy, toilets, and handwashing.** *Lancet* 2009, **374:**1032-1035.
5. Ngure FM, Reid BM, Humphrey JH, Mbuya MN, Pelto G, Stoltzfus RJ: **Water, sanitation, and hygiene (WASH), environmental enteropathy, nutrition, and**

early child development: making the links. *Ann N Y Acad Sci* 2014, **1308:**118-128.
6. Dangour AD, Watson L, Cumming O, Boisson S, Che Y, Velleman Y, Cavill S, Allen E, Uauy R: **Interventions to improve water quality and supply, sanitation and hygiene practices, and their effects on the nutritional status of children.** *Cochrane Database Syst Rev* 2013**:**CD009382.
7. Bowen A, Agboatwalla M, Luby S, Tobery T, Ayers T, Hoekstra RM: **Association between intensive handwashing promotion and child development in Karachi, Pakistan: a cluster randomized controlled trial.** *Arch Pediatr Adolesc Med* 2012, **166:**1037-1044.
8. Arnold BF, Null C, Luby SP, Unicomb L, Stewart CP, Dewey KG, Ahmed T, Ashraf S, Christensen G, Clasen T, et al: **Cluster-randomised controlled trials of individual and combined water, sanitation, hygiene and nutritional interventions in rural Bangladesh and Kenya: the WASH Benefits study design and rationale.** *BMJ Open* 2013, **3:**e003476.
9. **WASH Benefits Kenya Primary Outcome Analysis Plan Update** [https://osf.io/7urqa/]
10. Squires J, Bricker D, Potter L: **Revision of a parent-completed development screening tool: Ages and Stages Questionnaires.** *J Pediatr Psychol* 1997, **22:**313-328.
11. WHO Multicentre Growth Reference Study Group: **WHO Motor Development Study: windows of achievement for six gross motor development milestones.** *Acta Paediatr Suppl* 2006, **450:**86-95.
12. WHO Multicentre Growth Reference Study Group: **WHO Child Growth Standards for length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: Methods and development.** Geneva, Switzerland: World Health Organization; 2006.
13. Ballard T, Coates J, Swindale A, Deitchler M: **Household Hunger Scale: Indicator Definition and Measurement Guide.** Washington, DC: Food and Nutrition Technical Assistance (FANTA-II) Project, FHI 360; 2011.
14. Jewell NP, Emerson R: **Current Status Data: An Illustration with Data on Avalanche Victims.** In *Handbook of Survival Analysis.* Edited by Klein JP, van Houwelingen HC, Ibrahim JG, Scheike TH: Chapman and Hall; 2013: 391-412
15. Diamond ID, McDonald JW, Shah IH: **Proportional hazards models for current status data: application to the study of differentials in age at weaning in Pakistan.** *Demography* 1986, **23:**607-620.
16. Ayer M, Brunk HD, Ewing GM, Reid WT, Silverman E: **An empirical distribution function for sampling with incomplete information.** *Annals of Mathematical Statistics* 1955, **26:**641-647.
17. Wood S: *Generalized Additive Models: An Introduction with R. .* Taylor and Francis; 2006.
18. Walker SP, Chang SM, Powell CA, Simonoff E, Grantham-McGregor SM: **Early childhood stunting is associated with poor psychological functioning in late adolescence and effects are reduced by psychosocial stimulation.** *J Nutr* 2007, **137:**2464-2469.

19. Walker SP, Wachs TD, Grantham-McGregor S, Black MM, Nelson CA, Huffman SL, Baker-Henningham H, Chang SM, Hamadani JD, Lozoff B, et al: **Inequality in early childhood: risk and protective factors for early child development.** *Lancet* 2011, **378:**1325-1338.
20. Kariger P, Frongillo EA, Engle P, Britto PM, Sywulka SM, Menon P: **Indicators of family care for development for use in multicountry surveys.** *J Health Popul Nutr* 2012, **30:**472-486.
21. Kroenke K, Spitzer RL, Williams JB: **The PHQ-9: validity of a brief depression severity measure.** *J Gen Intern Med* 2001, **16:**606-613.
22. Wood AM, White IR, Thompson SG: **Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals.** *Clin Trials* 2004, **1:**368-376.
23. Hernan MA, Hernandez-Diaz S, Robins JM: **A structural approach to selection bias.** *Epidemiology* 2004, **15:**615-625.