WASH Benefits Bangladesh Primary Outcome Analysis Plan Update

5 Feb 2016

Background

The WASH Benefits study published a detailed design and analysis protocol for primary outcomes in 2013.[1] The protocol included nearly all details of the analysis plan for primary outcomes: diarrhea and length for age Z-scores (LAZ). Here, we pre-specify updates to the analysis plan for the primary outcomes in the Bangladesh trial. We have also included the secondary outcomes stunting and severe stunting in this analysis plan (stunting defined as LAZ < -2, severe stunting defined as LAZ < -3) because we plan to report them in the primary outcome paper for Bangladesh. Below, we have listed all changes from the 2013 analysis plan along with a rationale (the changes are minor). Please refer to the design and rationale protocol for specific details about outcome definitions, measurement methods, and hypothesis tests.[1]

Primary Analysis: changes and additions from the original protocol, along with rationale

- In the original protocol, we specified that we would calculate 95% confidence intervals using a nonparametric bootstrap for both the unadjusted and adjusted estimates of our parameters of interest. At the time we expected to need to use a nonparametric bootstrap to calculate confidence intervals in the adjusted analysis and wanted to keep a parallel structure across the unadjusted and adjusted analyses. Advances in the `tmle` R software package since the original protocol was published now readily implement influence curve-based standard errors and confidence intervals that account for clustered outcome data.[2] We will use influence curve-based standard errors to calculate 95% confidence intervals in the adjusted analysis (with clusters defining independent units). Therefore, we elected to use the paired t-test for unadjusted LAZ comparisons and the Mantel-Haenszel prevalence ratio and difference for unadjusted diarrhea and stunting comparisons,[3] with randomization block (groups of 8 village clusters) defining matched pairs/stratification. The confidence intervals based on the paired t-test, Mantel Haenszel estimator, and influence curve are asymptotically equivalent to the nonparametric bootstrap, but are computationally much simpler.
- In the original protocol, we specified that we would calculate one-sided *P*-values because we expected the interventions to only be beneficial. After a September 2014 collaborative meeting with investigators from the related SHINE trial (NCT01824940) and the published results from a study in Kenya that found iron fortification could increase enteric pathogen abundance in young children,[4] we decided to take a conservative approach and report two-sided *P*-values for all hypothesis tests.

- As specified in the original protocol, we will conduct the adjusted analyses using targeted maximum likelihood estimation (TMLE) on child-level outcomes. TMLE is implemented in the `tmle` package in the R statistical software.[2] TMLE enables the use of the SuperLearner[5] ensemble machine learning algorithm to flexibly adjust for the pre-specified covariates in our original protocol. We will include the following model library in the ensemble (`SuperLearner` R package arguments in parenthesis): the simple mean (`SL.mean`), main effects generalized linear models (`SL.glm`), generalized linear models with two-way interaction and stepwise selection using AIC (`SL.step.interaction`),[6] Bayesian generalized linear models (`SL.bayesglm`),[7] generalized additive models degree 2 (`SL.gam`),[8] and penalized elastic net regression (`SL.glmnet`).[9] We chose this model library because it enables flexible relationships between covariates and the outcome but does not include algorithms with internal bootstrapping or bagging (e.g., RandomForest[10]), which can lead to overfitting the data.
- We added child birth order and removed maternal literacy from our set of pre-specified adjustment covariates. A recent working paper based on Indian national survey data documented large differences in stunting based on birth order and sex preferences.[11] Maternal literacy was not measured in Bangladesh because our pilot study found extreme redundancy between literacy and education. We removed time between intervention delivery and measurement because this characteristic could not be measured in control clusters (no intervention) and because it was post-randomization. All maternal, household, and compound characteristics were measured at enrollment (before randomization) except for maternal height, which was measured at the 1 or 2 year follow-up visits to avoid the influence of pregnancy, since mothers were pregnant at enrollment. The final adjustment covariate list is below:
    - ID of the field staff member who recorded the measurement
    - Month of measurement, to account for seasonal variation
    - Child age (days)
    - Child sex
    - Child birth order
    - Mother's age (years)
    - Mother's height (cm)
    - Mother's education level (no education, primary, secondary)
    - Household food insecurity (4-level HFIAS categories)[12]
    - Number of children < 18 years in the household
    - Number of individuals living in the compound
    - Distance (in minutes) to the household's primary drinking water source
    - Housing materials (floor, walls, roof) and household assets
        - Assets measured. Has: electricity, wardrobe, table, chair or bench, watch or clock, khat, chouki, working radio, working black/white or color television, refrigerator, bicycle (not child's toy), motorcycle, sewing machine, mobile phone, land phone

- We will pre-screen covariates to assess whether they are associated with each outcome prior to including them in adjusted statistical models.[13] 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 < 0.2 in the adjusted analysis. We will also exclude any covariates that have very little variation in the study population (prevalence <5%).

Tests for potential between-cluster spillover effects (contamination), changes and additions from the original protocol, along with rationale

In Appendix 5 of the study's protocol, we outlined a nonparametric test for between-cluster spillover effects (i.e., contamination).[1] Although we expect such contamination to be very unlikely due to 1 km geographic buffers included between each study cluster in the Bangladesh trial, following our original protocol we will conduct formal tests to rule out such effects, which could bias the measures of intervention toward the null (by making outcomes across arms appear more similar than if arms were independent). Below, we have added details to the analysis plan for this statistical test that were not included in the original protocol.

- We will assess potential between-cluster spillovers for the following outcomes:
    - Primary outcomes:
        - Child diarrhea
        - Length-for-age
    - Behavioral indicators
        - Households have stored, chlorinated drinking water (measured by residual chlorine)
        - The participant reports that the youngest child's most recent defecation was either directly into the latrine or the feces were disposed of into the latrine
        - Households in the bari have a latrine with a functional water seal
        - Households have at least one handwashing station with soap and water present
        - The participant reports hearing any messages on infant/child nutrition and or Sonamoni (lipid based nutrient supplement)
- The original statistical test incorporated information about population density. However, we did not collect information about overall population density as part of the trial, so the models for between-cluster spillovers will not adjust for local population density.
- Our original analysis plan defined between-cluster spillovers as a function of the number of treated compounds within specific distances. We will estimate spillovers as a function of the distance to the nearest treated cluster instead. Our rationale is that a blinded interim analysis has shown that there is very little variation in the number of compounds within each cluster and the density of treatment does not vary greatly within each block. Thus, the distance to the nearest treated compound is likely to be as strong a predictor of spillovers as the treatment density, and the interpretation of spillovers associated with the distance to the nearest compound is more straightforward. This approach has been used in another assessment of spillovers in a randomized trials of interventions to prevent malaria.[14] We will estimate the association between each outcome (Y) among control compounds (T=C) and the distance to the nearest treated compound assigned to arm $T=t$, (D^T), adjusting for baseline covariates (X). We will fit the model $E[Y \mid T, X, D^T]$ nonparametrically using Super Learner ensemble machine learning and TMLE and will

include the same algorithm library specified above in the primary analysis. To test for the presence of between-cluster spillovers, we will use a clustered permutation test for each treatment T to test the following null hypothesis:

$$\psi = \mathrm{E}[Y \mid T=\mathrm{C}, X, D^T=d^T_{20}] - \mathrm{E}[Y \mid T=\mathrm{C}, X, D^T=d^T_{80}] = 0,$$

where d^T_{20} is the 20th percentile of the distribution of distance to the nearest treated compound for treatment $T=t$, and d^T_{80} is the 80th percentile. If between-cluster spillovers are present, we would expect this value to be greater than 0 for length-for-age Z-scores and less than 0 for diarrhea prevalence (i.e., compounds that are closer to treatment compounds should benefit from intervention more than compounds that are further away from treatment compounds). We are pre-specifying the 20th and 80th percentiles of the distance distribution because they are close to the lower and upper range of the distribution (i.e., close to the extremes), but still have sufficient observations to ensure that we can estimate the quantities with reasonable precision.

- In addition to distance, we will test whether spillovers occurred through the following shared social facilities:
    - Children's preschool where the respondent's children go or will go to school
    - Market the respondent goes to most often
    - Church/mosque/temple the household members attend
    - Health facility nearest to respondent
    - We define F as an indicator of shared social facilities. F=1 if the control compound shares a facility with a treatment compound. F=0 if the control compound does not share a facility with a treatment compound. To test for the presence of between-cluster spillovers through shared social facilities, we will use a clustered permutation test for each treatment T to test the following null hypothesis among control compounds:

      $$\psi = \mathrm{E}[Y \mid T=\mathrm{C}, X, F=1] - \mathrm{E}[Y \mid T=\mathrm{C}, X, F=0] = 0$$

      We will estimate this parameter using Super Learner ensemble machine learning and TMLE, as in the analysis of between-cluster spillovers based on distance to the nearest treated compound.

References

1. Arnold BF, Null C, Luby SP, 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(8):e003476.

2. Gruber S, van der Laan M. tmle: An R Package for Targeted Maximum Likelihood Estimation. J Stat Softw 2012;51(13):1–35.

3. Donner A, Klar N. Confidence interval construction for effect measures arising from cluster randomization trials. J Clin Epidemiol 1993;46(2):123–31.

4. Jaeggi T, Kortman GAM, Moretti D, et al. Iron fortification adversely affects the gut microbiome, increases pathogen abundance and induces intestinal inflammation in Kenyan infants. Gut 2015;64(5):731–42.

5. van der Laan MJ, Polley EC, Hubbard AE. Super Learner. Stat Appl Genet Mol Biol 2007;6(1):1544–6115.

6. Venables WN, Ripley BD. Modern Applied Statistics with S. 4th ed. Springer; 2002.

7. Gelman A, Jakulin A, Pittau MG, Su Y-S. A weakly informative default prior distribution for logistic and other regression models. Ann Appl Stat 2008;2(4):1360–83.

8. Hastie T, Tibshirani R. Generalized Additive Models. London: Chapman and Hall; 1990.

9. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw 2010;33(1):1–22.

10. Breiman L. Random Forests. Mach Learn 2001;45(1):5–32.

11. Jayachandran S, Pande R. Why Are Indian Children So Short? [Internet]. Cambridge, MA: National Bureau of Economic Research; 2015 [cited 2015 Oct 22]. Available from: http://www.nber.org/papers/w21036.pdf

12. Coates J, Swindale A, Bilinsky P. Household Food Insecurity Access Scale (HFIAS) for measurement of food access: indicator guide (v.3) [Internet]. FANTA; 2007. Available from: http://www.fantaproject.org/sites/default/files/resources/HFIAS_ENG_v3_Aug07.pdf

13. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. Stat Med 2002;21(19):2917–30.

14. Hawley WA, Phillips-Howard PA, Kuile FOT, et al. Community-Wide Effects of Permethrin-Treated Bed Nets on Child Mortality and Malaria Morbidity in Western Kenya. Am J Trop Med Hyg 2003;68(4 suppl):121–7.