Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

__Kiersten Derby_____ _04/11/13_____

[Student's name typed] Date

The relationship between inherited blood disorders and measures of iron status among young children in Kenya

By

Kiersten S. Derby

Degree to be awarded: MPH

Department of Epidemiology

_____ [Chair's signature]

Kevin M. Sullivan

Committee Chair

_____ [Member's signature]

Parmi S. Suchdev

Committee Member

The relationship between inherited blood disorders and measures of iron status among young children in Kenya

By

Kiersten S. Derby

B.S., Brown University, 2007

Emory Univeristy

2013

Faculty Thesis Advisor:

Parminder S. Suchdev, MD, MPH

Kevin M. Sullivan, PhD, MPH, MHA

Abstract

The relationship between inherited blood disorders and measures of iron status among young children in Kenya

By Kiersten S. Derby

Background: Inherited blood disorders may influence iron indicators, which would have implications in areas with high burden of disease.

Objective: To determine if there is an association between inherited blood disorders and iron biomarkers.

Design: We conducted a population-based, cross-sectional survey of 854 children aged 6-35 months in western Kenya. Participants were tested for sickle cell, α-thalassemia, and G6PD deficiency. Ferritin, transferrin receptor (TfR), and zinc protoporphyrin (ZP) were measured, and TfR/ferritin index was calculated. Linear regression, adjusting for sociodemographic characteristics, malaria, and inflammation, was used to assess the association between blood disorders and iron biomarkers.

Results: Inherited blood disorders were common; 18.7% had sickle cell disease or trait, 48.0% had abnormal α-thalassemia genotype, and 6.8% had G6PD deficiency. The percentage of the population with abnormal iron biomarkers varied by indicator, ranging from 19.2% according to ferritin < 12 ug/L to 97.8% using ZP > 80 μmol/mol. Mean unadjusted TfR was highest among children with HbSS genotype compared to HbAS and HbAA (ANOVA $p<0.0001$). Mean unadjusted ZP was higher among boys with normal genotype compared to those with G6PD deficiency ($p=0.02$). In multivariate analysis, G6PD deficiency was an independent predictor of ZP among boys (β-coefficient = -0.17, $p=0.04$). There was interaction between sickle cell and malaria ($p=0.01$); malaria was a predictor of ZP among children without sickle cell, but not a predictor of ZP among those with sickle cell.

Conclusions: In areas with high prevalence of inherited blood disorders, genotypic differences may independently affect iron biomarkers, particularly TfR and ZP.

The relationship between inherited blood disorders and measures of iron status among young children in Kenya

By

Kiersten S. Derby

B.S., Brown University, 2007

Emory Univeristy

2013

Faculty Thesis Advisor:

Parminder S. Suchdev, MD, MPH

Kevin M. Sullivan, PhD, MPH, MHA

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

In partial fulfillment of the requirements for the degree of

Master in Public Health

In Department of Epidemiology

2013

Table of Contents:

Thesis documents

Appendices

i. Exhaustive
ii. Reduced
iii. Final

Introduction

Iron deficiency is an important cause of anemia. Surveys performed in Kenya have found that over 65% of Kenyan children under 3 years of age are anemic, with a significant portion of these thought to be caused by iron-deficiency (1). However, laboratory assessment of iron status is difficult, as measures of iron status are affected by several factors including age, sex, infection/inflammation, and pregnancy (2). Furthermore, laboratory testing is expensive and invasive, and difficult to perform in low-resource settings (2). Newer non-invasive techniques to assess iron status are currently under development (2).

Biomarkers of iron status frequently measured in cross-sectional surveys include ferritin (an indicator of iron body stores), soluble transferrin receptor (TfR; an indicator of erythropoietic intensity and iron requirements), and zinc protoporphyrin (ZP; a measure of iron-deficient erythropoiesis attributable to low marrow iron stores). Calculations based on these measurements, including TfR/ferritin index, have also been used to estimate iron status in children since they are good estimates of body iron stores (3). The Joint WHO/CDC Technical Consultation on the Assessment of Iron Status at the Population Level recommends measuring the concentrations of hemoglobin for the assessment of anemia, the use of serum ferritin in areas where infectious diseases are less common, TfR in areas where inflammation is prevalent, and mention the usefulness of measuring one or more of the acute phase proteins (e.g., C-reactive protein [CRP] and alpha-glycoprotein [AGP]) (4). These recommendations reflect the fact that inflammation is known to influence the measurement of ferritin and zinc protoporphyrin, and, to a lesser extent, hemoglobin and TfR (5,6). The effect of inflammation on iron indicators is of particular concern in developing countries, where infections such as malaria and HIV are common (7). The use of a correction factor, based on two acute phase proteins biomarkers (CRP and AGP) to adjust iron status indicators for the effects of inflammation is one method to account for inflammation without excluding data (8). Adjusting for indicators of inflammation in regression models also helps to achieve this goal. Regression modeling also allows for adjustment to account for other potential factors associated with inflammation and iron biomarkers—such as demographics, inherited blood disorders, and malaria status—that the correction factor approach and exclusion approach do not address.

Inherited blood disorders are known to be common in sub-Saharan Africa (9,10). In Kenya, for example, previous studies have found that more than two-thirds of children in western Kenya have at least one measured blood disorder (including sickle cell disease or trait, α-thalassemia trait or disease, G6PD deficiency, or haptoglobin 2-2 genotype) (11). Few studies, however, have evaluated the independent association of inherited blood disorders with iron status indicators.

A better understanding of the relationship between inherited blood disorders and iron indicators may also be particularly important in areas with concurrent high prevalence of blood disorders and malarial infection, such as the region in Zanzibar in which the Pemba trial was conducted. This trial found that among children who were not iron deficient (defined as ZP<80 μmol/mol), there was an increase in severe adverse events (a composite of hospital admissions and deaths) among those who were supplemented with iron and folic acid, compared to placebo; iron deficient children who were

supplemented, however, had reduced rates of severe adverse events (12). These conclusions resulted in a change in WHO policy regarding universal iron supplementation in areas of high malaria burden (13).

The objectives of this study were 1) to use bivariate analysis to determine if the mean iron indicators (SF, TfR, ZP, and TfR/SF index) vary among Kenyan children with and without the inherited blood disorders sickle cell, α-thalassemia, or G6PD deficiency; and 2) to use multiple linear regression modeling to determine if the inherited blood disorders are independent predictors of the iron indicators, adjusting for identified confounders and interaction terms.

Subjects and Methods:

Study population and sample

The study was part of a larger longitudinal study, the Nyando Integrated Child Health and Education Project (NICHE), which evaluated the effectiveness of the promotion and sale of evidence-based health products, including micronutrient powders in 60 study villages during 2007–2010. Details of NICHE are described elsewhere (1,14,15).

In brief, this cross-sectional survey assessed 1,348 randomly selected children, aged 6-35 months, across 60 villages (30 intervention villages and 30 comparison villages) from the Nyando Division of western Kenya in August 2010. The probability of a village's selection was proportional to its size. Residents were primarily of Luo ethnicity, engaged in subsistence farming, and lived in compounds consisting of a single main house surrounded by one to three additional households. Using an updated 2010 household census that was conducted in the study area, 19 compounds were randomly selected per village. Lists of selected compounds were provided to the field team, and all children aged 6 to 35 months living in these compounds were eligible to participate. Written informed consent was obtained from all participating households. Trained fieldworkers used questionnaires to collect data from the mothers of the study participants on demographic and socio-economic factors, hygiene, sanitation, child feeding practices and child morbidity during the preceding 24h. Anthropometric measurements were also collected using standardized procedures. Children severely anemic (hemoglobin < 7.0 g/dL) or with clinical malaria (fever with positive malaria smear) were referred for treatment to the nearest hospital or clinic. Institutional review boards of the Kenya Medical Research Institute and the U.S. Centers for Disease Control and Prevention approved the study.

Of the children approached for enrolment, 1079 met criteria for enrolment and 197 were excluded before enrolment (33 refusal, 124 unavailable for enrollment, 40 other), leaving 882 children enrolled. The resulting response rate was 882/1079 (81.9%). Twenty-eight participants were excluded at time of analysis (21 missing hemoglobin, 3 out of age range, 4 no recorded iron indicator). The final sample population consisted of 854 children (**Figure 1**).

Assessment of health and nutrition status

Capillary blood was obtained from children by trained laboratory technicians through a finger stick for hemoglobin (Hb) measurement, malaria smear preparation, and Microtainer® collection. Details of the laboratory analyses are described in detail elsewhere (1,8). Hb was measured in the field using a HemoCue® B-Hemoglobin photometer (Ängelholm, Sweden). The instruments were calibrated daily. According to the WHO thresholds for children ages six months to five years, anemia was defined as Hb < 11.0 g/dL (16). Thick blood smears were prepared, stained with Giemsa, and observed using a light microscope by the KEMRI/CDC malaria lab in Kisian, Kenya. Approximately 500 µL of capillary blood was also collected into heparinized microcontainers and transported on ice to the project laboratory within 6 hours of collection. The remaining blood was centrifuged and the plasma separated, collected and stored at -40ºC. Samples were transported to Germany for analysis of ferritin, transferrin receptor (TfR), retinol binding protein (RBP), and C-reactive protein (CRP) using a sandwich enzyme-linked immunosorbent assay technique (17). The CDC laboratory oversaw the quality control and quality assurance of the specimen analysis.

Genotyping for HbS and the most common form of alpha-thalassemia in Africa, caused by a 3.7-kilobase pair delation of the alpha-globin chain, was performed by typed by polymerase chain reaction at the KEMRI-Wellcome Trust Laboratories in Kilifi, Kenya. Details of laboratory analyses are described elsewhere (8,18,19). Children who were heterozygous for the β^s mutation of the HBB gene were defined as having sickle cell trait, while homozygotes were defined as sickle cell disease. Children with a single α-globin deletion (-α/αα) were defined as heterozygotes or α-thalassemia silent carriers, while those with two α-globin deletions (-α/-α) were defined as homozygotes or α^+thalassaemia trait (19).

Statistical analysis

All statistical analysis was done using SAS 9.3 (SAS Institute Inc., Cary, NC). Significance was defined as $p<0.05$. The distributions of hemoglobin, ferritin, TfR, TfR/ferritin index, ZP, CRP, AGP, and RBP were assessed for normality, and all were found to be non-Gaussian (based upon a skewness and/or kurtosis value more extreme than -1 or 1), except for hemoglobin, AGP, and RBP. For variables with non-Gaussian distribution, log-transformation was performed. Values were back-transformed before presentation for ease of interpretation. Variables with non-Gaussian distributions were presented as geometric means and 95% confidence intervals (CI), while Gaussian distributions were presented as mean and 95% CI. Categorical variables were presented as proportion and 95% CI.

The thresholds for defining abnormal values for the previously mentioned biochemical indicators are as follows: serum ferritin (SF), <12 µg/L; TfR, >8.3 mg/L; ZP>80 µmol/mol; CRP, >5 mg/L; AGP, >1.0 g/L; and RBP, <0.70 µmol/L (20). The TfR/ferritin index was defined as the value of TfR divided by the value of

ferritin, with an index value of >500 defined as abnormal (21). Inflammation was categorized as any inflammation or no inflammation. Any inflammation was defined as either CRP >5 mg/L or AGP >1 g/L (or both), while no inflammation was defined as CRP ≤5 mg/L and AGP ≤1 g/L, 2).

Sickle cell was categorized as sickle cell trait (HbAS) or disease (HbSS) compared to a reference of a normal hemoglobin genotype (HbAA). α-thalassemia was taken as a three-level variable: normal (αα/αα), heterozyote (αα/α-), and homozygote (-α/-α). G6PD activity was categorized as normal or deficient.

For the demographic and socioeconomic variables, age was measured in months, low socioeconomic status was defined as socioeconomic status quintiles 1-2, low maternal education was defined as less than completed primary education, recent tea consumption was defined as consumption in previous 24 hours, and recent Sprinkles use was defined as consumption in previous 24 hours. Sprinkles are single-serving sachets of powdered vitamins and minerals that may be mixed into any semisolid food before consumption without altering the food's taste or color (1). According to the WHO Child Growth Standards, stunting was defined as a height-to-age Z-score less than -2; wasting was defined as a weight-to-height Z-score of less than -2; and underweight was defined as a weight-to-age Z-score of less than -2.

Baseline characteristics of the population were analyzed using PROC SURVEYFREQ for categorical analyses and PROC UNIVARIATE for continuous analyses. Univariate analysis was conducted for each inherited blood disorder. The proportion of children with iron indicators beyond the threshold values was calculated for each level of the inherited blood disorder and the chi-squared test (Student's t-test) was used for comparison of proportions. The mean for each iron indicator was also calculated for each level of the inherited blood disorder and analysis of variance (ANOVA) was used for comparison of means (using PROC GLM).

Multiple linear regression analysis using PROC SURVEYREG was chosen to model the biological effects of the independent variables (inherited blood disorders) on the dependent variables (iron indicators), accounting for cluster design. A total of six multivariate models were created, including a model for sickle cell and each of the four measures of iron status, as well as the findings in ANOVA univariate analysis that reached or approached statistical significance: α-thalassemia and TfR, and ZP and G6PD deficiency among boys. In each model, the iron indicator was taken as the dependent variable and the inherited blood disorder was the primary independent variable. We evaluated and confirmed linearity of continuous variables by analysis of scatter plots. Two-way interaction was tested by adding an interaction term into the linear regression model with only the primary variable and additional variable of interest. The two-way interaction term was included in the exhaustive model if the p-value was less than 0.05. If the p-value was greater than or equal to 0.05, the interaction term was excluded from the exhaustive model. Confounding was then assessed. Confounding was determined to be present if the crude parameter estimate for the inherited blood disorder from the simple linear regression model differed by greater than 10% from the adjusted parameter estimate from the multiple linear regression model with the iron indicator as outcome, the inherited blood disorder as the primary variable, and the

potential confounder as the additional variable. The independent variables that were evaluated for interaction and confounding were age, sex, inflammation, RBP, stunting, wasting, underweight, α-thalassemia, G6PD deficiency, tea consumption in past 24 hours, Sprinkle use in past 24 hours, presence of malaria parasitemia, fever in past 24 hours, socioeconomic status, and maternal education. For each model, an exhaustive model was first created, including all interaction terms and confounders. We then attempted to identify a more parsimonious model by removing the least significant independent variable and assessing whether the regression coefficient for the inherited blood disorder was changed by more than 10%. If it was, the variable was retained in the model; otherwise, the variable was dropped from the model. Any interaction term was dropped in the reduced model if the p-value was greater than 0.01. To arrive at the final model, we reran the reduced model in the presence of all of the blood disorders, including those that did not remain significant independent predictors, to adjust for all identified blood disorders.

In order to isolate the unique effect of the primary iron indicator variable that was not explained by the other measures of iron status, we also reran the models by forcing the other iron indicators (besides the iron indicator of interest in each particular model) into the exhaustive model, then proceeding with the above described modeling method.

For the statistically significant interaction terms, we evaluated the mean iron indicator for each subgroup using PROC SURVEYMEANS, which accounted for the cluster design of the survey.

Results:

Descriptive statistics of the 854 children from Nyando District, Kenya in 2010 are presented in **Table 1**. Inherited blood disorders were common in this population. A total of 492 children (57.6% of the population) had at least one of the blood disorders (sickle cell, α-thalassemia, or G6PD deficiency). The most prevalent blood disorder was α-thalassemia, with 9.6% being homozygous or trait (-α/-α) and 38.4% being heterozygous or silent carriers for the disease. Seventeen percent of the population had sickle cell trait, while 1.6% had sickle cell disease.

Using cut-off values described previously, the percentage of the population with abnormal iron biomarkers varied by iron indicator and ranged from 19.2% (95% CI: 15.8-22.7%) according to SF < 12 ug/L, to 97.8% (95% CI: 96.5-99.0%) using ZP > 80 µmol/mol. Additionally, 71.5% of the population was anemic (Hb<11.0 g/dL). Many children were found to have inflammation and/or infection, with 61.5% having any inflammation (defined as an elevated CRP, AGP, or both), 32.4% with malaria parasitemia, and 41.8% reporting a fever in the past 24 hours. Anthropometric measures indicated that 29.6% of the population was stunted, 12.1% was underweight, and 3.5% was wasted.

In univariate analysis, each of the four iron indicators were investigated, and means and proportions were presented, stratifying by inherited blood disorder status (**Table 2**). The mean TfR was significantly

higher among the HbSS genotype compared to either HbAS or HbAA (ANOVA p-value <0.0001); the other iron indicators did not vary significantly across sickle cell genotype. Mean TfR across α-thalassemia genotype approached but did not reach statistical significance (ANOVA p-value = 0.07). Mean ZP was significantly higher among boys with normal G6PD function compared to boys with G6PD deficiency (p=0.02).

Multiple linear regression modeling was used to isolate the effect of an inherited blood disorder on the iron indicator of interest by adjusting for interaction and confounding. We created four adjusted models that took sickle cell as the primary independent variable, and, consecutively, each of the four indicators of iron status as the dependent variable (**Table 3**). In Model 1, where log-transformed ferritin was the dependent variable, sickle cell was a borderline significant independent predictor of log-transformed ferritin (p=0.05). The other statistically significant predictors of log-transformed ferritin included malaria parasitemia, inflammation, fever in the past 24 hours, and the interaction term between sickle cell and socioeconomic status. Thirty five percent of the model's variance was described by the included variables (R^2 = 0.35). The significant interaction term indicates that among those from low SES households, children with sickle cell have a statistically significantly higher mean ferritin compared to the children without sickle cell (low SES children with sickle cell: mean ferritin 44.7 ug/L, 95%CI: 34.1-58.7; low SES children without sickle cell: mean ferritin 32.8 ug/L, 95% CI: 27.8-38.8; p=0.03). Among those from high SES households, children with sickle cell have a marginally significant lower mean ferritin compared to children without sickle cell (high SES children with sickle cell: mean SF 26.1 ug/L, 95% CI 20.4-33.4; high SES children with no sickle cell: mean ferritin 33.9 ug/L, 95% CI: 29.3-39.1; p=0.05). The biological plausibility of this interaction term may be argued.

For Model 2, the statistically significant predictors of log ZP included malaria parasitemia, inflammation, and the interaction term for sickle cell and malaria parasitemia; however, the variables in the model only explained 14% of the variance in log ZP (R^2=0.14). According to the interaction term, among those with malaria, children with sickle cell have a significantly lower mean ZP compared to children without sickle cell (malaria positive children with sickle cell: mean ZP 224.40 µmol/mol, 95% CI 200.4-251.3; malaria positive children without sickle cell: mean ZP 287.1 µmol/mol, 95% CI: 271.0-304.2; p=0.01). Among those without malaria, there is no statistically significant difference in mean ZP levels (malaria negative children with sickle cell: mean ZP 206.16 µmol/mol, 95% CI 179.9-236.2; malaria negative children without sickle cell: mean ZP 195.2 µmol/mol, 95% CI: 183.5-207.6;p=0.42). In other words, sickle cell is a significant predictor of mean ZP, but only among those with malaria.

The variables in model 3 only explained 14% of the variance in log TfR. The statistically significant predictors of log TfR included malaria parasitemia and inflammation, as well as stunting. α-thalassemia trait was a borderline significant predictor of log TfR (p=0.05). To interpret the β-coefficient on the log-scale, we used the following formula: back-transformed β-coefficient (expressed as percent change) = 100*(e^(log-transformed β-coefficient)-1). Therefore, a child with α-thalassemia trait is predicted to have a TfR that is 8.33% higher than the TfR of a child without α-thalassemia.

In model 4, infection and inflammation as measured by malaria parasitemia, any inflammation, and reported fever in past 24 hours were all significant predictors of log TfR/SF index; as were the demographic characteristics of age, underweight, and stunting, and the interaction term for sickle cell and underweight. According to the interaction term, sickle cell is a predictor of TfR/SF index among children who are underweight (HbAS or HbSS: mean TfR/SF index 672.8, 95%CI: 37.7-1265.6; HbAA: mean TfR/SF index 272.9, 95% CI: 197.7-376.7), but not among children who are not underweight (HbAS or HbSS: mean TfR/SF index 383.6, 95% CI 320.5-459.2; HbAA: mean TfR/SF index 392.7, 95% CI 351.1-439.3).

Table 4 presents the final parsimonious linear regression model for the dependent variable log TfR and the primary exposure variable of α-thalassemia (labeled model 5). α-thalassemia trait approaches significance as an independent predictor of log TfR (p=0.06), while α-thalassemia disease is not a statistically significant predictor (p=0.72). According to this model, children with α-thalassemia trait have an 8.33% increase in TfR compared to children with a normal genotype, although this finding does not reach statistical significance.

Model 6 (also shown in Table 4) investigates the association between the primary exposure G6PD deficiency on the dependent variable log ZP, adjusting for confounders, among boys. G6PD deficiency among boys was a significant independent predictor of log ZP (p=0.04). The other statistically significant predictors of log ZP included RBP, malaria parasitemia, and age. To interpret the β-coefficient on the log-scale, we again used the following formula: back-transformed β-coefficient (expressed as percent change) = 100*(e^(log-transformed β-coefficient)-1). According to this model, ZP among boys who are G6PD deficient is 15.63% less than a boy with normal G6PD function. The exposure variables in this model only accounted for 13% of the variance in ZP.

To isolate the unique effect of the primary variable on the iron indicator that was not explained by the other measures of iron status, we also reran the models by forcing the other iron indicators (besides the iron indicator of interest in each particular model) into the model (**Table 5,6**). According to these models, sickle cell was a statistically significant predictor of log-transformed ZP (p=0.03) and a borderline statistically significant predictor of log-transformed ferritin (p=0.05) and log TfR (p=0.06). Alpha-thalassemia remained a borderline significant predictor of log TfR (p=0.06). Because, with the exception of rerun model 1, two-way interaction terms fell out of the models due to non-significance, one may interpret the direction of the beta-coefficients in rerun models 2-5. In rerun model 2, a child with sickle cell disease or trait is predicted to have a ZP that is 7.69% less than a child with a normal hemoglobin genotype (Table 5). According to re-run model 3, a child with sickle cell disease or trait is predicted to have a ZP that is 8.33% higher than a child with a normal hemoglobin genotype, although this result does not achieve statistical significance (p=0.06).

Discussion

Based on our analysis of this large, population-based, cross-sectional survey of children aged 6-35 months in Nyando District, Kenya, genotypic differences in inherited blood disorders may affect iron biomarkers, particularly ZP. This is one of the only studies to evaluate the association of blood disorders and iron biomarkers in a resource-poor setting, where malaria and subclinical infection are prevalent. The need for such studies was reinforced by a working group participating in the Biomarkers of Nutrition for Development (BOND) project when they identified a need for research on the utility of iron biomarkers in populations with high prevalence of α-thalassemia and relevant genetic polymorphisms (2).

In our adjusted models, we found that G6PD deficiency among boys is a statistically significant predictor of ZP after adjustment for malaria parasitemia, inflammation, and other demographic characteristics. A boy with G6PD deficiency is predicted to have a 15.63% decrease in ZP, compared to a boy with normal G6PD activity. Previous research suggests that G6PD deficiency may be protective against severe malaria in hemizygous males (but not heterozygous females), particularly in the A- form of G6PD deficiency that is widespread in Africa (22). However, our results indicate that G6PD deficiency is an independent predictor of ZP even after controlling for malaria parasitemia, suggesting that this finding is not completely explained by malarial infection. The mechanism by which G6PD deficiency affects ZP, a measure of iron deficient erythropoiesis attributable to low iron supply in the bone marrow, is not known.

We also found that α-thalassemia trait is an independent predictor of elevated log TfR, although this finding did not reach statistical significance (p=0.06). Several previous studies have found a significant association between log TfR and α-thalassemia using multiple linear regression analysis. One such study, conducted in a population of 181 children in Vanuatu, found that in a model of log TfR against age, sex, α-globin genotype, and log ferritin, α-thalassemia trait and disease (as well as log ferritin) were both significant predictors of log TfR (23). Another study, conducted in Cambodia where genetic hemoglobin disorders are also prevalent, again found that alpha-thalassemia trait was a significant predictor of elevated log TfR (24). This association may be explained by the hypothesis that α-thalassemia causes ineffective erythropoiesis, leading to an elevated TfR, a measure of erythropoietic intensity (23). Additionally, it is interesting to note that in our study, malaria parasitemia and inflammation were statistically significant predictors of TfR in both model 3 and model 5, which is traditionally believed to not to be influenced by inflammation (4). George *et al.* also found that log TfR was significantly elevated by chronic inflammation (24). Grant *et al.* recently recommended that TfR was the best single biomarker for estimating the prevalence of iron deficiency in preschoolers (compared to ferritin, ZP, and the TfR/ferritin index), based upon the best kappa statistic for agreement with the multiple-criteria model. However, due to the mounting evidence that alpha-thalassemia may independently influence TfR, as well as a possible persistent association with inflammation, in populations known to have a high burden of inherited blood disorders, this recommendation should be used with caution (25).

Sickle cell (HbAA or HbAS) is a borderline statistically significant predictor of log ferritin (p=0.05), after adjustment for interaction terms and confounders. It is well established that ferritin, an acute phase reactant, is significantly impacted by age, sex, and infection and inflammation, among other things, but little is known of the effect of hemoglobin mutations on this iron indicator (4). To our knowledge, one previous survey has used multiple linear regression modeling taking the iron indicators as dependent variables and sickle cell as the primary exposure. In this study, Nyakeriga et al. (2005) found that HbAS genotype was negatively associated with log ferritin, adjusted for other factors including age, sex, fever, malaria parasitemia, and inflammation measured by CRP (β=-0.20, p=0.037) (26). This study also found that neither α-thalassemia heterozygosity nor α-thalassemia homozygosity were significantly associated with log ferritin. George *et al*. studied another hemoglobin genotype, hemoglobin E variant, and found that hemoglobin E trait (with or without α-thalassemia trait) and hemoglobin E disease were significant predictors of increased ferritin (24).

While sickle cell is not a statistically significant predictor of ZP, the significant interaction term between sickle cell and malaria parasitemia may shed some light on the association between inherited blood disorders, inflammation, and iron biomarkers. We found that among children with malaria, mean ZP in children with sickle cell was significantly lower than the mean ZP in children without sickle cell. However, among children without malaria, there was no statistically significant difference in mean ZP levels. Therefore, sickle cell is a significant predictor of mean ZP only among children with malaria.

Despite the uncertainty about the preferred measures of iron status, important policy decisions have been made based upon research that may not accurately assess iron status in the field, since the gold standard of bone marrow biopsy is not feasible. In the Pemba trial, Sazawal *et al*. found that all-cause mortality and hospital admission were 12% higher among children receiving iron and folic acid supplementation compared to placebo, and in a substudy analysis stratified by iron deficiency that children who were not iron deficient had a non-statistically significant increase in severe adverse events, compared to the placebo arm, while iron deficient children were found to have a reduction of severe adverse events (12). As a result of this trial, the World Health Organization and the United Nations Children's Fund advised that in regions with a high prevalence of malaria, iron and folic acid supplementation should only be administered to those who have been identified as anemic (12,27). In light of our analysis, there could be genotype differences (e.g., G6PD deficiency) between the Pemba study arms that may have been reflected in the ZP values, which therefore affected the reported malaria incidence. Of note, a similar trial to the Pemba trial conducted in a low malaria prevalence region in Nepal used serum ferritin to measure iron status and found no difference in mortality between children supplemented with iron and folic acid compared to placebo (28).

There are several limitations of this study. First, results of this study can only be applied to preschool children in the Nyando district and cannot be generalized to other parts of sub-Saharan Africa. Second, because of the cross-sectional design, our conclusions are limited to correlation and not causal relationships; a longitudinal, cohort study would be needed to demonstrate causation. Additionally, our R^2 values were low, ranging from 0.14-0.35 for the models without the inclusion of all the iron

indicators, suggesting that there is significant variance in the outcome variable that is not explained by the model. The R^2 values did increase dramatically with the addition of the other iron indicators into the models, ranging from 0.41-0.60; these values approximated those found by George *et al.*, who also included additional iron indicators in their models. Finally, we did not measure all iron status indicators (including hepcidin and red blood cell indices), and we did not do bone marrow biopsies, the gold standard for measuring iron status. Despite these limitations, our findings remain important because of the large number of children sampled and the comprehensive measurement of three iron indicators (plus one calculated index), two markers or inflammation, the presence of malaria parasitemia, and three inherited blood disorders.

In conclusion, we have shown that the inherited blood disorders of sickle cell, α-thalassemia, and G6PD deficiency all may affect measures of iron status, thus complicating the assessment of iron status in population with a high prevalence of blood disorders. However, sickle cell is not a significant predictor of log index, both before and after adjustment for the other iron indicators; because the index was not affected by the inherited blood disorders, this may be the best choice for measuring iron status in areas with a high burden of inherited blood disorders. A greater understanding of the association between inherited blood disorders and iron indicators, as well as the biological mechanisms of these relationships, is needed.

Future Directions

I will continue to analyze this data before submission of the manuscript for publication. The next step will be to re-run the models without including the supplemental iron indicators, but assessing if each of the blood disorders is a significant confounder or two-way interaction term (like was done for other variables in above analysis). If a blood disorder is a significant confounder or two-way interaction term, if will be included in the exhaustive model; if it is not a significant confounder or two-way interaction term, it will not be included in the exhaustive model (unlike above, where we forced the blood disorders into the model).

In creating a reduced model from the exhaustive model, the first step will be to remove the least significant interaction term one at a time and exclude the interaction term if the beta coefficient of the primary variable *or* of the interaction term changes by less than 10%. It the beta coefficients change by greater than 10%, they will remain in the model. Once only statistically significant interactions are in the model ($p<0.05$), the least significant variables that were not in an interaction term will be removed one at a time and confounding assessed (as described above) until only significant and/or important confounders remain in the model.

I will then use this same approach described above, this time including the iron indicators in the model, starting with evaluating whether they are significant confounders or two-way interaction terms. This will provide me with a second version of the models, one that accounts for the other iron indicators and

isolates the effect the inherited blood disorder on the primary iron indicator of interest. I will also look at Pearson correlation coefficients between the three iron markers—if they are strong, this would supplement the argument of not including all of the iron markers in the models. There is also statistical support of including all of the iron markers in the models, as described above, and a published precedent as all previous studies with similar methods have included all of their measured iron markers in their models. I will also include a supplemental description in the revised methods section about how each iron biomarker measures different components of iron metabolism.

References

1. Suchdev PS, Ruth LJ, Woodruff BA, Mbakaya C, Mandava U, Flores-Ayala R, Jefferds MED, Quick R: Selling Sprinkles micronutrient powder reduces anemia, iron deficiency, and vitamin A deficiency in young children in Western Kenya: a cluster-randomized controlled trial. Am J Clin Nutr 2012,95(5):1223-30.

2. Raiten DJ, Namaste S, Brabin B, Combs G, Jr., L'Abbe MR, Wasantwisut E, Darnton-Hill I: Executive summary--Biomarkers of Nutrition for Development: Building a Consensus. Am J Clin Nutr 2011, 94(2):633S-650S.

3. Lynch S: Improving the assessment of iron status. Am J Clin Nutr 2011, 93:1188-9.

4. WHO/CDC: Assessing the iron status of populations: Report of a Joint World Health Organization/Centers for Disease Control and Prevention Technical Consultation on the Assessment of Iron Status at the Population Level. In. Geneva, Switzerland; 2007: 1-35.

5. Kung'u JK, Wright VJ, Haji HJ, Ramsan M, Goodman D, Tielsch JM, Bickle QD, Raynes JG, Stoltzfus RJ: Adjusting for the acute phase response is essential to interpret iron status indicators among young Zanzibari children prone to chronic malaria and helminth infections. J Nutr 2009, 139(11):2124-2131.

6. Wieringa FT, Dijkhuizen MA, West CE, Northrop-Clewes CA, Muhila l: Estimation of the effect of the acute phase response on indicators of micronutrient status in Indonesian infants. J Clin Pathol 2002, 132:3061-3066.

7. Northrop-Clewes CA: Interpreting indicators of iron status during an acute phase response--lessons from malaria and human immunodeficiency virus. Ann Clin Biochem 2008, 45(Pt 1):18-32.

8. Grant FK, Suchdev PS, Flores-Ayala R, Cole CR, Ramakrishnan U, Ruth LJ, Martorell R: Correcting for inflammation changes estimates of iron deficiency among rural Kenyan preschool children. J Nutr 2012, 142(1):105-111.

9. Weatherall DJ: Hemoglobinopathies worldwide: present and future. Current Molecular Medicine 2008, 8:592–9.

10. Modell B, Darlison M: Global epidemiology of haemoglobin disorders and derived service indicators. Bulletin of the World Health Organization 2008, 86:480–487.

11. Suchdev PS, Ruth LJ, Earley M, Macharia A, Williams TN: The burden and consequences of inherited blood disorders among young children in western Kenya. Matern Child Nutr 2012. Epub ahead of print.

12. Sazawal S, Black RE, Ramsan M, Chwaya HM, Stoltzfus RJ, Dutta A, Dhingra U, Kabole I, Deb S, Othman MK, et al. Effects of routine prophylactic supplementation with iron and folic acid on admission to hospital and mortality in preschool children in a high malaria transmission setting: Community-based, randomised, placebo-controlled trial. Lancet 2006, 367(9505): 133-143.

13. World Health Organization. (2007a). Conclusions and recommendations of the WHO Consultation on prevention and control of iron deficiency in infants and young children in malaria-endemic areas. Food Nutr Bull 2007a, 28(4 Suppl):S621-S627.

14. CDC, 2007. Baseline Data from the Nyando Integrated Child Health and Education Project—Kenya. MMWR Morb Mortal Wkly Rep 56: 1109–1113.

15. Suchdev PS, Ruth LJ, Obure A, Were V, Ochieng C, Ogange L, Owuor M, Ngure F, Quick R, Juliao P: Monitoring the marketing, distribution, and use of Sprinkles micronutrient powders in rural western Kenya. Food Nutr Bull 2010,31: 168–178.

16. WHO: Iron Deficiency Anemia: Assessment, Prevention, and Control--A Guide for Program Managers. In.; 2001.

17. Erhardt JG, Estes JE, Pfeiffer CM, Biesalski HK, Craft NE: Combined measurements of ferritin, soluble transferrin receptor, retinol binding protein, and C-reactive protein by an inexpensive, sensitive, and simple sandwich enzyme-linked immunosorbent assay technique. J Nutr 2004, 134:3127-3132.

18. Williams TN, Wambua S, Uyoga S, Macharia A, Mwacharo JK, Newton CRJC, et al. Both heterozygous and homozygous α+ thalassemias protect against severe and fatal Plasmodium falciparum malaria on the coast of Kenya. Blood 2005;106(1):368-371.

19. Waterfall CM, Cobb BD. Single tube genotyping of sickle cell anaemia using PCR-based SNP analysis. Nucleic acids research 2001;29(23):e119.

20. Thurnham DI, McCabe LD, Haldar S, Wieringa FT, Northrop-Clewes CA, McCabe GP: Adjusting plasma ferritin concentrations to remove the effects of subclinical inflammation in the assessment of iron deficiency: a meta-analysis. Am J Clin Nutr 2010, 92(3):546-555.

21. Akesson A, Bjellerup P, Berglund M, Bremme K, Vahter M: Soluble transferrin receptor: longitudinal assessment from pregnancy to postlactation. Obstet Gynecol 2002, 99(2):260-266.

22. Guido A, Fairhurst RM, Doumbo OK, Wellems TE, Diallo DA. X-linked G6PD deficiency protects hemizygous males but not heterozygous females against severe malaria. PLoS Med 2007, 4(3):e66.

23. Rees DC, Williams TN, Maitland K, Clegg JB, Weatherall DJ. Alpha thalassaemia is associated with increased soluble transferrin receptor levels. Br J Haematol. 1998;103:365–9.

24. George J, Yiannakis M, Main B, Devenish R, Anderson C, An US, Williams SM, Gibson RS: Genetic hemoglobin disorders, infection, and deficiencies of iron and vitamin A determine anemia in young Cambodian children. J Nutr 2012, 142(4):781-7.

25. Grant FK, Martorell R, Flores-Ayala R, Cole CR, Ruth LJ, Ramakrishnan U, Suchdev PS. Comparison of indicators of iron deficiency in Kenyan children. AJCN 2012, 95(5):1231-7.

26. Nyakeriga AM, Troye-Blomberg M, Mwacharo JK, Wambua S, Williams TN. Nutritional iron status in children with α+ thalassemia and the sickle cell trait in a malaria endemic area on the coast of Kenya. Haematologica 2005, 90:552-4.

27. Iron and Malaria Technical Working Group. Considerations for the safe and effective use of iron interventions in areas of malaria burden: Full technical report. Ed: Raiten D, Namasté S, Brabin B. September 2009.

28. Tielsch J, Khatry S, Stoltzfus R, Katz J, LeClerq S, Adhikari R, Mullany LC, Shresta S, Black RE. Effect of routine prophylactic supplementation with iron and folic acid on preschool child mortality in southern Nepal: Community-based, cluster-randomised, placebo-controlled trial. Lancet 2006, 367(9505): 144-152.

Figure 1: Inclusion Criteria

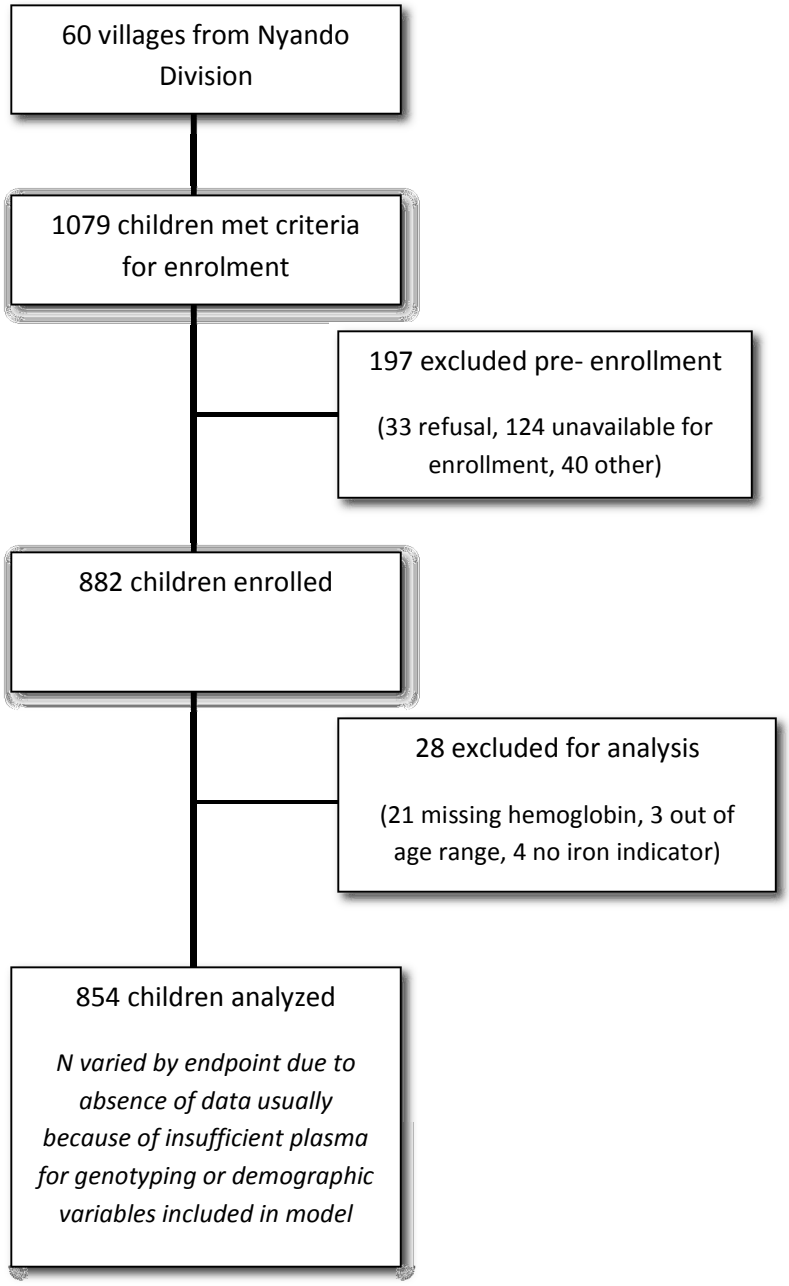


Table 1: Characteristics of enrolled children aged 6-35 months in Nyando District, Kenya, 2010 (N=854)

Continuous Variables	N	Mean (95% CI)[1]
Age (months)	854	21.5 (20.9-22.0)
Ferritin (µg/L)	847	33.0 (30.6-35.6)
ZP (µmol/mol)	853	219.0 (211.2-227.1)
TfR (mg/L)	847	12.7 (12.3-13.2)
TfR / Ferritin index	847	385.3 (354.6-418.7)
Hb (g/dL)	854	9.6 (9.5-9.8)
CRP (mg/L)	847	9.0 (8.2-9.9)
AGP (g/L)	847	1.2 (1.1-1.2)
RBP (µmol/L)	847	0.9 (0.8-0.9)

Categorical Variables	n/N[2]	Percentage (95% CI)
Ferritin <12 ug/L	163/847	19.2 (15.8-22.7)
ZP >80 µmol/mol	834/853	97.8 (96.5-99.0)
TfR >8.3 mg/L.	635/847	75.0 (71.3-78.7)
Index >500	334/847	39.4 (36.2-42.6)
Anemia (Hb<11.0 g/dL)	611/854	71.5 (68.0-75.1)
CRP >5 mg/mL	289/847	34.1 (29.6-38.6)
AGP >1 g/L	515/847	60.8 (56.1-65.5)
Any inflammation[3]	525/854	61.5 (56.8-66.1)
RBP < 0.70 µmol/L	269/854	31.5 (27.6-35.4)
Hemoglobin type		
HbAA	693/853	81.2 (78.2-84.2)
HbAS	146/853	17.1 (14.3-19.9)
HbSS	14/853	1.6 (0.8-2.5)
α -Thalassemia genotype		
Normal (αα/αα)	427/822	51.9 (48.3-55.6)
Heterozygote (-α/αα)	316/822	38.4 (35.2-41.6)
Homozygote (-α/-α)	79/822	9.6 (7.6-11.7)
G6PD genotype		
Normal	769/825	93.2 (91.0-95.4)
Deficient	56/825	6.8 (4.6-9.0)
Male sex	429/854	50.2 (46.8-53.7)
Stunted (HAZ <-2)	252/850	29.6 (26.5-32.8)
Wasted (WHZ <-2)	30/851	3.5 (1.9-5.2)
Underweight (WAZ <-2)	103/852	12.1 (9.7-14.4)
Malaria parasitemia	275/848	32.4 (28.3-36.5)
Fever in past 24 hours	344/823	41.8 (38.0-45.6)
Low SES[4]	334/837	39.9 (34.5-45.4)
Low Maternal Education[5]	395/830	47.6 (43.1-52.1)
Tea consumption in past 24 hours	687/830	82.8 (79.6-85.9)
Sprinkles use in past 24 hours[6]	91/831	11.0 (8.2-13.7)

[1]Non-gaussian continuous variables presented as geometric mean and 95% confidence interval (anti-log). Age, AGP, hemoglobin and RBP are normally distributed thus presented as mean and 95% confidence interval. Categorical variables presented as proportion and 95% Wald confidence limits .

[2]n represents the numerator; N represents the total sample size or denominator.

[3]Inflammation defined as any of the following: elevated CRP with normal AGP, elevated CRP and AGP, or elevated AGP with normal CRP.

[4]Defined as SES quintiles 1-2.

[5]Low Maternal Education defined as less than completed primary education.

[6]Sprinkles are individual sachets of micronutrient powders.

Abbreviations: AGP, alpha-glycoprotein; CRP, C-reactive protein; HAZ, height-for-age z score; Hb, hemoglobin; RBP, retinol binding protein; SES, socio-economic status; TfR, soluble transferrin receptor; WAZ, weight-for-age z score; WHZ, weight-for-age z score; ZP, zinc protoporphyrin.

Table 2: Univariate analysis of four measures of iron status by inherited blood disorder among enrolled children aged 6-35 months in Nyando District, Kenya, 2010 (N=854)

	Measures of Iron Status																			
	Serum ferritin (SF)					Zinc protoporphyrin (ZP)					Soluble transferrin receptor (TfR)					TfR/SF Index				
	N	SF <12 ug/L [n (%)]	Chi-squared p-value	Mean (SD)	ANOVA p-value	N	ZP>80 umol/mol [n (%)]	Chi-squared p-value	Mean (SD)	ANOVA p-value	N	TfR >8.3 mg/L [n (%)]	Chi-squared p-value	Mean (SD)	ANOVA p-value	N	Index >500 [n (%)]	Chi-squared p-value	Mean (SD)	ANOVA p-value
Hemogloblin Type																				
HbAA	449	120 (26.7)		59.2 (66.6)		692	677 (97.8)		255.6 (145.3)		688	512 (74.4)		14.5 (8.6)		688	269 (39.1)		885.9 (2072.4)	
HbAS	98	31 (31.6)	0.24	54.9 (63.6)	0.36	146	142 (97.3)	0.78	243.4 (142.0)	0.64	144	110 (76.4)	0.57	15.1 (9.7)	<0.0001	144	60 (41.7)	0.60	1021.3 (1917.8)	0.61
HbSS	11	1 (9.1)		80.4 (66.0)		14	14 (100)		260.1 (129.9)		14	12 (85.7)		26.8 (21.5)		14	4 (28.6)		534.0 (484.0)	
α -Thalassemia genotype																				
Normal	279	76 (27.2)		58.4 (67.6)		426	422 (99.1)		246.6 (138.9)		426	308 (72.3)		14.1 (8.3)		426	166 (39.0)		896.8 (2211.8)	
Heterozygote	203	56 (27.6)	0.99	60.4 (65.6)	0.86	316	304 (96.2)	0.02	260.8 (150.1)	0.42	311	242 (77.8)	0.23	15.7 (10.6)	0.07	311	124 (39.9)	0.96	924.0 (1982.1)	0.98
Homozygote	53	15 (28.3)		56.4 (59.1)		79	76 (96.2)		251.8 (150.7)		78	59 (75.6)		14.8 (8.6)		78	30 (38.5)		900.0 (1429.0)	
G6PD genotype																				
Males																				
Normal	256	83 (32.4)	0.39	56.7 (65.8)	0.50	379	371 (97.9)	0.39	278.1 (155.1)	0.02	377	298 (79.1)	0.20	16.2 (10.2)	0.11	377	173 (45.9)	0.43	1125.4 (2463.3)	0.22
Deficient	25	6 (24.0)		48.9 (53.3)		34	34 (100.0)		213.8 (108.5)		34	30 (88.2)		13.4 (6.0)		34	18 (52.9)		609.2 (510.5)	
Females																				
Normal	247	56 (22.7)	0.91	60.6 (66.5)	0.81	389	380 (97.7)	0.51	233.1 (130.8)	0.53	386	271 (70.2)	0.84	13.7 (8.4)	0.86	386	124 (32.1)	0.39	736.5 (619.6)	0.74
Deficient	14	3 (21.4)		64.1 (60.6)		22	21 (95.5)		251.5 (180.4)		22	15 (68.2)		14.0 (8.6)		22	9 (40.9)		619.6 (771.6)	

N represents the total sample size or denominator; n represents the numerator.

Analysis does not account for cluster survey design.

Table 3: Multiple linear regression analysis with log ferritin, log ZP, and log TfR as dependent variable among all enrolled children aged 6-35 months in Nyando District, Kenya, 2010

	Model 1: Log ferritin		Model 2: Log ZP		Model 3: Log TfR		Model 4: Log Index	
n	747		788		781		750	
R^2	0.35		0.14		0.14		0.18	
	β Coefficient	*p-value*	*β Coefficient*	*p-value*	*β Coefficient*	*p-value*	*β Coefficient*	*p-value*
Primary Exposure								
Sickle cell disease or trait (reference: normal genotype)	-0.22	0.05	0.05	0.42	0.05	0.38	-0.10	0.30
Blood disorders								
G6PD deficiency (reference: normal genotype)	0.09	0.42	-0.10	0.17	-0.05	0.41	-0.15	0.32
α-Thalassemia								
Heterozygous vs Normal	0.01	0.88	0.04	0.35	0.08	0.05	0.08	0.42
Homozygous vs Normal	0.13	0.28	0.02	0.75	0.06	0.36	-0.04	0.78
Morbidity and inflammation								
Malaria parasitemia	0.61	<0.0001	0.32	<0.0001	0.26	<0.0001	-0.35	0.002
Inflammation[1]	0.89	<0.0001	0.17	0.0003	0.19	<0.0001	-0.68	<0.0001
Fever in past 24 hours	0.27	0.001					-0.25	0.01
Child demographics and characteristics								
Age							-0.01	0.003
Underweight (WAZ < -2)	0.11	0.32					-0.52	0.003
Stunted (HAZ < -2)					0.12	0.01	0.42	0.001
Low socioeconomic status (reference: quintiles 3-5)	-0.12	0.10					-0.08	0.33
Sprinkles use in past 24 hours	0.06	0.57						
RBP								
Interaction term								
Sickle cell * Socioeconomic status	0.59	0.003						
Sickle cell * Malaria parasitemia			-0.29	0.01				
Sickle cell * Underweight							0.92	0.01

[1]Inflammation defined as any stage of inflammation (incubation, early convalescence, late convalescence) and reference is no inflammation.

Abbreviations: HAZ, height-for-age z score; RBP, retinol binding protein; TfR, transferrin receptor; WAZ, weight-for-age z score; ZP, zinc protoporphyrin.

Table 4: Two multiple linear regression models with an inherited blood disorder as the independent variable and an iron indicator as the dependent variables among all enrolled children or limited to only boys aged 6-35 months in Nyando District, Kenya, 2010

	Model 5: α-Thalassemia and log TfR		Model 6: G6PD deficiency and log ZP among boys	
n	744		375 [1]	
R^2	0.16		0.13	
	β coefficient	*p-value*	*β-coefficient*	*p-value*
Primary Exposure				
α-Thalassemia				
Heterozygous vs Normal	0.08	0.06		
Homozygous vs Normal	0.02	0.72		
G6PD deficiency (reference: normal genotype)			-0.17	0.04
Blood disorders				
α-Thalassemia				
Heterozygous vs Normal			0.02	0.8
Homozygous vs Normal			-0.06	0.64
G6PD deficiency (reference: normal genotype)	-0.06	0.43		
Sickle cell disease or trait (reference: normal genotype)	0.05	0.30	-0.17	0.04
Morbidity and inflammation				
Malaria parasitemia	0.24*	<0.0001	0.21	0.01
Inflammation[2]	0.20*	<0.0001	0.09	0.21
Fever in past 24 hours	0.33	0.33	0.08	0.19
RBP			-0.25	0.03
Child demographics and characteristics				
Age			-0.01	0.03
Female sex (reference: male sex)	-0.15*	<0.0001		
Tea consumed in past 24 hours	0.07	0.10		
Sprinkles consumed in past 24 hours			0.08	0.41
Underweight (WAZ <-2)			0.01	0.93
Stunted (HAZ < -2)	0.09*	0.04		

[1] Analysis limited to boys

[2] Inflammation defined as any stage of inflammation (incubation, early convalescence, late convalescence) and reference is no inflammation.

Abbreviations: HAZ, height-for-age z score; RBP, retinol binding protein; WAZ, weight-for-age z score.

Table 5: Multiple linear regression analysis with log ferritin, log ZP, log TfR, and log TfR/SF index as dependent variable, including all iron indicators, among all enrolled children aged 6-35 months in Nyando District, Kenya, 2010

	Rerun model 1: Log ferritin		Rerun model 2: Log ZP		Rerun model 3: Log TfR		Rerun model 4: Log TfR/SF Index	
n	753		784		780		750	
R^2	0.41		0.60		0.60		0.42	
	β Coefficient	*p-value*	*β Coefficient*	*p-value*	*β Coefficient*	*p-value*	*β Coefficient*	*p-value*
Primary Exposure								
Sickle cell disease or trait (reference: normal genotype)	-0.21	0.05	-0.08	0.03	0.08	0.06	0.10	0.25
Blood disorders								
G6PD deficiency (reference: normal genotype)	0.02	0.8	-0.06	0.3	0.02	0.69	0.01	0.96
α-Thalassemia								
Heterozygous vs Normal	0.04	0.58	-0.01	0.59	0.05	0.06	0.02	0.81
Homozygous vs Normal	0.13	0.24	-0.01	0.71	0.05	0.24	-0.07	0.56
Iron status indicators								
Log ferritin			-0.07	<0.0001	-0.01	0.38		
Log ZP	-0.46	<0.0001			0.71	<0.0001	1.22	<0.0001
Log TfR	-0.10	0.27	0.72	<0.0001				
Morbidity and inflammation								
Malaria parasitemia	0.74	<0.0001	0.08	0.04	0.12	0.01	-0.65	<.0001
Inflammation[1]	0.99	<0.0001	0.07	0.04	0.11	0.002	-0.90	<.0001
Fever in past 24 hours	0.31	<0.0001					-0.34	<.0001
Child demographics and characteristics								
Low socioeconomic status (reference: quintiles 3-5)	-0.14	0.04					0.01	0.91
Sprinkles use in past 24 hours								
RBP			-0.21	<0.0001	0.19	0.001		
Age							-0.01	0.06
Underweight (WAZ < -2)	0.15	0.20					-0.30	0.07
Stunted (HAZ < -2)					0.05	0.03	0.28	0.004
Interaction term								
Sickle cell * Socioeconomic status	0.51	0.01						

[1]Inflammation defined as any stage of inflammation (incubation, early convalescence, late convalescence) and reference is no inflammation.

Abbreviations: HAZ, height-for-age z score; RBP, retinol binding protein; TfR, transferrin receptor; WAZ, weight-for-age z score; ZP, zinc protoporphyrin.

Table 6: Two multiple linear regression models with an inherited blood disorder as the independent variable and an iron indicator as the dependent variables, rerun to include all iron indicators in analysis, among all enrolled children or limited to only boys aged 6-35 months in Nyando District, Kenya, 2010

	Rerun Model 5: α-thalassemia and log TfR		Rerun Model 6: G6PD and log ZP among boys[1]	
n	784		377	
R^2	0.59		0.56	
	β Coefficient	*p-value*	*β Coefficient*	*p-value*
Primary Exposure				
α-Thalassemia	0.05	0.05		
G6PD deficiency (reference: normal genotype)			-0.09	0.12
Blood disorders				
α-Thalassemia				
Heterozygous vs Normal			-0.03	0.50
Homozygous vs Normal			-0.07	0.24
G6PD deficiency (reference: normal genotype)	0.03	0.58		
Sickle cell disease or trait (reference: normal genotype)	0.08	0.04	-0.13	0.02
Iron status indicators				
Log ferritin	-0.02	0.23	-0.08	0.0001
Log ZP	0.70	<0.0001	0.66	<0.0001
Morbidity and inflammation				
Malaria parasitemia	0.10	0.03	0.16	0.01
Inflammation[2]	0.09	0.01		
Fever in past 24 hours			0.06	0.17
RBP			-0.34	<0.0001

[1] Analysis limited to boys.

[2] Inflammation defined as any stage of inflammation (incubation, early convalescence, late convalescence) and reference is no inflammation.

Abbreviations: HAZ, height-for-age z score; RBP, retinol binding protein; ZP, zinc protoporphyrin.

Appendix 1: Questionnaire

FOLLOW-UP HOUSEHOLD QUESTIONNAIRE

TEAM CODE: _____ INTERVIEWER CODE: _____ TODAY'S DATE: _____/_____/2010

HOUSEHOLD – DEMOGRAPHICS

The household questionnaire should be completed by an adult living in the selected household.

H1. SUBLOCATION **SublocID** (CIRCLE ONE)	01-Achego 02-Ahero 03- Ayucha 04- Ayweyo 05- Border 1 06- Border 2 07- Kobongo 08- Kakmie 09- Katolo 10-Kochogo Central 11-Kochogo North 12-Kochogo south 13-Magina 14-Nyakongo 15-Ombaka 16-Wanganga
H2. VILLAGE **Villagename**	----------
H3. CLUSTER NUMBER **Cluster** *(enter from cluster listing form)*	☐☐
H4. **NYING WUON DALA EN NG'A?** NAME OF THE COMPOUND HEAD **DalaName**	----------
H5. DALA NUMBER **DalaNumber** *(enter from cluster listing form)*	☐☐
H6. HOUSEHOLD ID **HHID**	☐☐ ☐☐ ☐☐ ☐☐ subloc cluster dala # HH #
H7. **NYINGI EN NG'A?** RESPONDENT'S NAME **RespName**	----------
H8. **HIKI ADI ?** RESPONDENT'S AGE **Rage**	☐☐ years
H9. **RESPONDENT'S SEX** **Rsex**	Male (**wuoyi)** ..... 1 Female(**nyako)** 2
H10. **OD NI MARU KOSO UPANGO?** **OWNRENT** ARE YOU TENANTS IN THIS HOUSE OR IS IT OWNED BY THE FAMILY?	Owned (**ot mari**) ..... 1 Rented (**ikodesa**) 2

H11. **OT KA RUM ADI MA JI NINDE?** **RoomNum** HOW MANY ROOMS IN THE HOUSE ARE USED FOR SLEEPING?	_____ Rooms (**rums)**
H12. **UN GI STIMA E ODU KA?** **Electricity** IS THERE ELECTRICITY IN THIS HOUSE?	No (**ooyo)** ..... 0 Yes (**eeh)** ..... 1 Don't know (**ok ang'eyo)**..... 99

H13. **BE UN GI:** DO YOU CURRENTLY HAVE ANY OF THE FOLLOWING IN YOUR HOUSE?
(Read. Mark all that apply)

Item	No **(ooyo)**= 0 Yes **(eeh)**= 1
NYAKALONDO (RADIO) **Radio**	0 1
TELEBISEN (TELEVISION) **TV**	0 1
FRIJ (REFRIGERATOR) **Refrig**	0 1
NDIGA (BICYCLE) **Bike**	0 1
PIKIPIKI (MOTORCYCLE) **Piki**	0 1
MATOKA (A CAR) **Car**	0 1
SIMB JOPOSTA (LANDLINE TELEPHONE) **TelLand**	0 1
SIMB ONG'WE YAMO (MOBILE PHONE) **TelCell**	0 1
JATICH MONDIKI (A HOUSEHELP) **DomWork**	0 1

Household SWAP Module

H14. **BENDE IN KATA JAODNI MORO EN JAUSO MAR SWAP?** **Vendor** ARE YOU OR ANYONE IN YOUR HOUSEHOLD A SWAP VENDOR?	No(**podi**)……...0 Yes(**ase ngiew'o**)……1	
H15. **BENDE JAUS GIGE SWAP/NICHE OSEBIRO E ODU KA?** **SwapVisit** HAS ANY VENDOR VISITED YOUR HOUSE TO SELL HEALTH PRODUCTS?	No(**podi**)……...0 Yes(**osebiro**)……1 Don't know (**ok ang'eyo**)……99	**IF NO OR DK, GO TO H18**
H16. **BENDE NING'IEWO GIR SWAP/NICHE MORO AMORA?** **BuySWAP** DID YOU BUY ANY HEALTH PRODUCTS?	No(**podi**)……...0 Yes(**ase ngiew'o**)……1 Don't know (**ok ang'eyo**)……99	**IF NO or DK, GO TO H18**

H17. **ANG'O MANING'IEWO?** DID YOU BUY? *(Read. Mark all that apply)*	WaterGuard(**waterguard)** **BuySWAPWG**.....0 / 1 PUR (**PUR**) **BuySWAPPUR**.....0 / 1 Modified Clay Pot(**agulu molos man gi fereji)**…0 / 1 Bednets (ITN) (**net mar suna)** .....0 / 1 Condoms(**kondom)**…**BuySWAPcon**.....0 / 1 Sprinkles… **BuySWAPSpr**.....0 / 1 Fortified Flour(**mogo mayom) BuySwapFlow** 0 / 1 Soap(**sabun)BuySwapSoap**.....0 / 1 Savlon(**yath mar savlon)BuySwapSav**.....0 / 1 Other(**moro mopogore)BuySwapOth**.....0 / 1 Don't know (**ok ang'eyo)BuySwapDK**.....0 / 1

WATER & HYGIENE MODULE **Read: "Now we would like to talk with you about the water you use in your home"**	
H18. **PI MA UMODHO E OT KAE KAWUONO UYUDO KOA KANYE? HHSRC** WHAT DRINKING WATER SOURCE ARE YOU USING TODAY? *(Don't read. Mark only one)*	Pond (**Dago**), River (**Aora**), Dam / Earthpan (**Yawo**), or Lake (**Nam**) .....1 Borehole **(Kisima mokuny gi masin)** .....2 Rain water catchment **(Pii koth)**.....3 Covered Well **(Kisima manigi pump)** .....4 Open Well **(Kisima maonge pump)**.....5 Spring **(Soko moger)** .....6 Piped Water **(Pii fereji)**.....7 Water vendors **(Jo us pii)** .....8 From school **(skul)**.....9

	Other **moro mopogore** ……..……..88 Don't know(**ok ang'eyo)**……..……..99	
H19. **BENDE NITIE GIMA UTIMO NE PI MONDO OBED MABER MAR MODHO? WATSAFE** DO YOU DO ANYTHING TO THE WATER TO MAKE IT SAFE FOR DRINKING?	No (**da)**…….....0 Yes**(nitie**)……..……1 Don't know (**ok ang'eyo)**……..……99	**IF NO OR DK, GO TO H21**
H20. **ANG'O MAITIMONE?** WHAT DO YOU DO TO IT? (DON'T READ. MARK ALL THAT APPLY)	Use WaterGuard (**atiyo gi waterguard)**…. ……..... 1 Boil water (**chwako pii)** ……. 1 Filter water (**a chungo pii)** ……..…… 1 Use PuR (**atiyo gi PUR)**..... 1 Use Aluminum sulphate- (**atiyo gi Aluminium)**…. 1 Other **(moro mopogore)** …….. 1	
H21. **BENDE UKANO PI MODHO? Store** DO YOU STORE DRINKING WATER?	No**(ok wa kan**)…….....0 Yes**(wakano**)……..……1	**IF NO, GO TO H23**
H22. **UKANO PI MODHONO E ANG'O? StoreWat** WHERE DO YOU STORE THE DRINKING WATER? *(DON'T READ. MARK ONLY ONE)*	Plastic jerrycan(**kube mar plastic)** ……..……1 Buckets(**ndoo)** ……..……2 Ordinary clay pot(**agulu)** ……..…3 Improved clay pot (narrow mouth with tap) (**agulu moketi e tap)**……..……4 Barrel (**pipa/daram mar pii)** ……..……5 Do not store drinking water.....6 Other **moro mopogore**……..……..88	
H23. **BENDE ISEWINJO WATERGUARD? HearWG** HAVE YOU HEARD ABOUT WATER GUARD?	No (**podi)**……..……...0 Yes (**asewinjo)**……..……1 Don't know (**ok ang'eyo)**……..……99	**IF NO OR DK, GO TO H32**

H24. **NIWINJE KOA KANYE?** (If Yes) WHERE DID YOU HEAR ABOUT IT? *(Don't read. Mark all that apply)*	Radio (**redio)**… …….. 1 Newspaper (**gaset)** ….……..... 1 My child in school (**nyathina manie skul)** …..... 1 Brochure/Poster (**kalatas mondiki mar lendo** .. 1 WaterGuard t-shirt **(T-shat mar WaterGuard)**… 1 Community Resource Persons (**jogo matiyo e gweng')**…….……..……..…. 1 Promotion show(**tuke mag lendo)** ……..……. 1 Community meetings/chiefs baraza (**chokruok/barasa)**…….……... 1 CARE Kenya (**jo CARE Kenya)**…….. 1 Wall painting(**picha mar korot maduong')**….... 1 Health facility (**kar thieth)**..…… 1 Neighbor / family / friends (**jogo ma wadak go/osiepe)** …… 1 Health Officer/Nurse (**jathieth/sista matiyo e hospital)** …….……..……..…. 1 SWAP/NICHE… ……..…… 1 Other **(moro mopogore)**… ……..…….. 1
H25. **WATER GUARD MAROMO NADE MA ITIYOGO E LITA 20 MAR PI MALER?** HOW MUCH WATER GUARD DO YOU USE	One capful(**wi chupa achiel)**……..……1 Other **(moro mopogore)** ……..……88

TO TREAT 20LITERS OF CLEAN WATER? **WGClear** *(DON'T READ. MARK ONLY ONE)*	Don't know (**ok ang'eyo)** …….. 99	
H26. **WATERGUARD MAROMO NADE MA ITIYOGO E LITA 20 MAR PI MA OLIL?** HOW MUCH WATER GUARD DO YOU USE TO TREAT 20L of DIRTY WATER? **WGTurb** *(DON'T READ. MARK ONLY ONE)*	Two capfuls(**wi chupa ariyo)**……1 Don't have or use turbid water (**ok ati gi pii dago/molil)** ……..……2 Other **(moro mopogore)**……..……88 Don't know (**ok ang'eyo)**…….. 99	
H27. **KA ISETHIEDHO PIGI GI WATERGUARD OBER MAR MODHO BANG' SECHE ADI?** AFTER HOW LONG IS THE WATER TREATED WITH WATERGUARD SAFE FOR DRINKING? **WGWait**	Less than 20 minutes **(matin ne dakika 20)** ……1 20 minutes or more **(dakika 20 kata mokalo)** ….2 Don't know **(ok ang'eyo)**……..……99	
H28. **BENDE ISEGATHIEDHO PIGI GI WATERGUARD?** HAVE YOU EVER TREATED YOUR WATER WITH WATER GUARD? **WGEverTrt**	No (**podi)**……..0 Yes (**asethiedhe)**……..……1 Don't know **(ok ang'eyo)** …….. 99	**IF NO OR DK, GO TO H30**
H29. **PI MA UMODHO SANI BENDE OTHIEDH GI WATERGUARD? WGCurTrt** IS THE WATER YOU ARE DRINKING CURRENTLY TREATED WITH WATER GUARD**?**	No (**ok othiedhe)**…….0 Yes (**othiedhe)**……..……..1 Don't know **(ok ang'eyo)**…….. 99	**IF YES OR DK, GO TO H31**
H30. (IF NO) **ANG'O MOMIYO?** WHY IS THAT? *(DON'T READ. MARK ALL THAT APPLY)*	Expensive(**beche tek**) ……..1 Bad taste/smell (**ok omit/dum marach)** ……..1 It resembles jik (**ochal gi jik)** ……1	All responss go to H31

	Don't need (**ok adwar)** ……..… 1 Too difficult to use (**otek tiyo go)** ……..... 1 Don't know where to buy it (**ok ang'eyo kuma ing'iewe)** ……. 1 Other **(moro mopogore)** ….……1 Don't know (**ok ang'eyo)** ……1
H31. **SANI BENDE IN GI SABUN EI OT KA?** DO YOU CURRENTLY HAVE SOAP IN THE HOUSE? **Soap**	No (**onge)**……..……0 Yes (**an go)**…..……1 Don't know **(ok ang'eyo)**…… 99
H32. **UTIYO GI CHOO MANE?** WHAT TOILET FACILITY DO YOU USE? **Toilet** *(DON'T READ. MARK ONLY ONE)*	In the bush or on the ground (**e bungu kata laro)**1 Latrine(**choo mokuny)** ……..……2 Flush toilet(**choo mantie e ot)** ……..……3 River(**aora)** ……..……..4 Other **(moro mopogore)** ……..……88

HH – OBSERVATIONS	
H33. WHAT TYPE OF **ROOFING** DOES THIS HOUSE HAVE? **Roof**	Thatch (**lum)**……..……1 Iron sheet(**mabati)** ……..……2 Tile/Asbestos sheets (**tail miketo e wi ot) ….**…3 Wood *(**bao)***……..…..4 Cement (**simiti)** ……5 Other **(moro mopogore)** ……..……88
H34. WHAT IS THE **FLOORING** MATERIAL? **FLOOR**	Dung/Mud (**owuoyo/loo)** ……..……1 Metal **(chuma)** ……..……..……2 Wood (**bao)**……..……..……3 Cement(**simiti)** ……..……4 Tile/Linoleum **(tail)** ……5 Other **moro mopogore** ……88

H35. WHAT IS THE MATERIAL USED FOR THE **WALLS**? **WALL**	Dung/Mud (**owuoyo/loo)** ……..……1 Metal(**chuma)** ……..…2 Wood(**bao)** ……..…3 Cement/Plaster(**simiti)** ……4 Bricks/blocks/stones(**matafari/kite)..**……5 Other **moro mopogore** ……..88	
H36. **BENDE ANYALO NENO GI MA IKANO E PII MAR MODHO?** MAY I SEE YOUR DRINKING WATER CONTAINER? **ObsStore**	Plastic jerrycan(**kube mar juala)** ……..……..…..1 Buckets(**ndoo)**……..…… .....2 Ordinary clay pot(**agulu)** ……..…….…...3 Improved clay pot (narrow mouth with tap) (**agulu man gi tap)** ……..……4 Barrel(**pipa/daram)** ……....……5 Container not present(**gir pii ong'e)** ……..……6 Refused (**otamore)** ……..……77 Other **(moro mopogore)** ……..……88	**IF REFUSE or not present, GO TO H39**
H37. *Confirm presence of **lid**.* **ObsLid**	No (**onge)**……..……...0 Yes (**nitie**)……..……1	
H38. *Test drinking water* **ObsChlor**	Negative (clear) (**ler)**……..…….... 0 Positive (pink) **(ratong'**)…….. 1 No water in the container(**pii onge E kube)**……..…2	
H39. **KELE WATERGUARD MA INGODO ANEE?** CAN I SEE YOUR WATERGUARD? **ObsWG**	Absent (**onge)**……..…….. 0 Present(**nitie)**……..…….. 1 Refused **(otamore)**……..… 77	
H40. **BENDE ANYALO NENO KALENDANI MAR SPRINKLES?** May I see your Sprinkles calendar? **ObsCal**	Absent (onge)…….. 0 Present ……..……1 Refused (otamore)…….…….. 77	

HHID | | | | | | | | |

MOTHER OF CHILD QUESTIONNAIRE

MOTHER DEMOGRAPHICS

The household questionnaire should be completed by the mother or caretaker for each child 6-35 months of age from each selected household.

M1. **NYING MAMA** MOTHER'S NAME	_____
M2. **HIK MAMA** **MOMAGE** MOTHER'S AGE	☐☐ Years
M3. **ICHOPO E OKANG' MANE MAR SOMO?** WHAT IS YOUR HIGHEST LEVEL OF EDUCATION **MomEduc**	None (**Onge**) ……1 Some Primary School **(Ok otieko primari skul)**……2 Completed Primary **(Otieko primary)**……....3 Some Secondary School **(Ok otieko secondary)**…..4 Completed Secondary School **(Otieko secondary)**...5 Any Trade School or University **(Skul mamoko kata mbalariany)**……..6 Other (**Mamoko**)……..……88 Don't know (**Akia**) …… 99
M4**. BENDE JOODI NE NITIERE NONRO MAR JO NICHE MANE ILIMO JI BANG' JUMBE ARIYO?** DID YOUR HOUSEHOLD PARTICIPATE IN THE NICHE STUDY WHERE PEOPLE VISITED THE HOUSE APPROXIMATELY EVERY TWO WEEKS? **NICHEHH**	No, never……0 Yes……..……..1 Don't know……..99

MOTHER SPRINKLES **Koro wadwaro wuoyo e wi gimachielo** "Now we would like to talk with you about a different subject."		
M5. **BENDE ISEWINJO KATA NENO GIMA ILUONGO NI 'SPRINKLES'?** HAVE YOU EVER HEARD OF SPRINKLES? **HearSP** ***(Show sachet of Sprinkles)***	No **(Podi)** …… … ……...0 Yes (**Eee**) …… … ……..……1	**IF NO, GO TO M7**
M6. **NIWINJO 'SPRINKLES' NI KANYE?** DID YOU HEAR ABOUT SPRINKLES FROM? ***(Read and mark each one yes or no)***	Martha/Cliff at training **SPTrn**…… 0 / 1 NICHE enumerators **SPEnum** ……..0 / 1 My child from school **(Nyathina mani e skul)**…….0 / 1 Community Health Worker (**Jopuonj mag gweng'**) ...0 /	

	1 Chiefs baraza (**Barasa mar gweng'**)……..…0 / 1 Church Leaders/at Church **(Jopuonj mar Kanisa/ e Kanisa)** ……**SpChurch**……..0 / 1 Health facility **(Kar thieth)** **SPFacil**..……..0 / 1 Neighbor / family / friends (**Jirani/watni/osiepeni)**.. 0 / 1 Health Officer/Nurse **(Ja helth/sista/jothieth mantiere e gweng')** …**SPHO**…….……0 / 1 Vendors **(Jous gige SWAP/NICHE) SPSwap**….. 0 / 1 Other **(Mamoko)** …**SPOth**……0 / 1 Don't know (**Akia)** …**SPDK**……0 / 1
M7. **ANG'O MABIRO E PACHI MOKUONGO KALUWORE GI SPRINKLES?** WHAT IS YOUR IMMEDIATE FIRST REACTION TO SPRINKLES? **SPRxn** *(Don't read. Mark only one)*	It's a good idea (**en paro maber)** ….. ……..1 It's a bad idea (**ok en paro maber)** …….. .....2 I am not sure **(ok an ga diera)** .....3 Don't know (**Akia)** ……..…...99
M8. **IPARO NI 'SPRINKLES' NI ITIYO GODO E YORE MAGE?** WHAT DO YOU THINK SPRINKLES IS USED FOR? *(Don't read. Mark all that apply)*	Appetizer (**Ndhandhu /keto dhok mamit)RxnApp**….1 Give energy, make active **(Medo teko) RxnEnergy** ..……1 Make child, family happy (**Keto nyathi, joot bedo gi mor)** …**RxnHappy**……1 Make child playful (**Keto nyathi hero tugo/ njejore)** …… RxnPlay……1 Grow healthy, make child healthy (**Miyo nyathi dongo kendo bedo kod ngima)** …**RxnHealth**……..1 Improved immunity (**Geng'o/kedo gi tuoche)** **RxnImmun**……..1 Prevent low blood, adds blood (**Medo remo)** .,,,,,,,,,,,,,.., 1 Make child stronger (**Keto nyathi bedo ma ratego**) … 1 Child smarter, build brain (**Nyathi bedo gi obuongo ma otegno / riek)** ……**RxnSmart**……1 Increase vitamin/minerals in body (**Medo chumbe mag**

	del) ……**RxnVit**……..……..……1 Sleep well/peacefully (**Nindo mayom/maber) ……..**1 Smooth healthy skin, prevent rashes (**Pien del bedo mayom, ma onge guonyo guonyo)** ….**RxnSkin**…….1 Hair strong, healthy, black (**Yier wich man gi ngima, ma otegno)**……**RxnHair**……1 Prevent diarrhea (**Geng'o diep)** …**RxnDiarr**……..…..1 Prevent malaria (**Geng'o malaria/midusi) RxnMal**…...1 Improve body development (**Keto del dongo maber) RxnDevel**……..……..……1 Other (**Mamoko**) **SpUseOth**…… ……...1 Don't know (**Akia) SpUseDK**…… ……..……1
M9. **'SPRINKLES' EN ANG'O?** WHAT ARE SPRINKLES? **SPWhat** *(Don't read. Mark only one)*	Powder with vitamins & minerals (or no mention of content) **(Poda man gi ndhandhu/chumbe mag del)** ……..……..……1 Drug (medicine, drug in powder form) (**Yath/Yien)**……..2 Food (e.g., fruits) (**Chiemo)** ……..…..3 Food supplement (might mention nutrients, food groups, v&m) **(Gik ma miyo chiemo teko mamoko)** ……4 Other **(Mamoko)**……88 Don't know **(Akia)**……..99
M10. **SPRINKLES IMIYO JOK MA HIKGI ADI?** WHAT AGE GROUPS ARE SPRINKLES MEANT FOR? **SPAge** *(Don't read. Mark only one)*	6 months to 5 years **(Dweche 6 nyaka higni 5)**….. .. .1 Under 5 years (Explicitly includes those less than 6 months)**(Ma hikgi tin ne 5)** ..… ……2 Young children (no age group mentioned) **(Nyithindo matindo)** ……... …. ……...3 Everybody **(Ng'ato ang'ata)** …… ……..4 Other **(Mamoko)**……..88 Don't know **(Akia)**……99
M11. **SPRINKLES ONEGO TIGO DIDI, TO MAROMO NADI?** **SPFreq**	1 sachet per day per child…….……1 2 sachet per week ……..…..2 1 sachet at every meal, every day ……..……...3 Episodic …….. ..……..4 1 sachet a week ……..……..5

HOW OFTEN SHOULD SPRINKLES BE USED? *(Don't read. Mark only one)*	Other **(Mamoko)**……..…88 Don't know **(Akia)**……..…99
M12. **CHIEMO MAROMO NADI MONEGO MEDIE SPRINKLES?** TO WHAT SIZE PORTION OF FOOD SHOULD YOU ADD SPRINKLES? **SPPortion** *(Don't read. Mark only one)*	Small portion a child can consume ……..…1 Other **(Mamoko)**……..88 Don't know **(Akia)**……99
M13. **OWINJORE IMI CHIEMO MOKETIE SPRINKLES THUOLO MAROMO NADI ?** HOW SOON AFTER ADDING SPRINKLES TO FOOD SHOULD YOU WAIT TO SERVE IT TO THE CHILD? **SPSoon** *(Don't read. Mark only one)*	Immediately serve to child **(sano sano)** …… ....1 Other **(Mamoko)**……..88 Don't know **(Akia)**……99
M14. **BENDE OWINJORE IKET SPRINKLES EI CHIEMO KAPOD CHIEK?** IS IT RECOMMENDED TO POUR IN THE SPRINKLES SACHET WHILE THE FOOD IS COOKING ON THE FIRE? **SPFire** *(Don't read. Mark only one)*	No **(Ooyo)** ……..…0 Yes **(Eee)** …… …… ……1 Don't know **(Akia)**……99
M15. **BENDE OWINJORE IMED SPRINKLES EI CHIEMO MALIW, KAKA PII, CHAK KATA**	No **(Ooyo)** ……..…0 Yes **(Eee)** …… …… ……1

CHAE? IS IT RECOMMENDED TO ADD SPRINKLES TO LIQUIDS? **SPLiq** *(Don't read. Mark only one)*	Don't know **(Akia)**……..99
M16. **GIN RANYISI MAGE MANYISO NI SPRINKLES TIYO?** WHAT ARE SIGNS THAT SPRINKLES IS WORKING? *(Don't read, mark all that apply)*	Increased appetite **(Medo dhok mamit)**.**AppSP**……..1 Increased energy **(Medo teko)** …**EnergSP**……1 Dark stool or change in color **(Losruok marateng')….**1 Loose stool, diarrhea **(Losruok marep rep, diep)**……1 Child happy **(nyathi mamor)**……**HappySP**……1 Child playful **(Nyathi mohero tugo/ma njejre)**…… 1 Child stronger **(Nyathi motegno)**….**StrongSP**……1 Child healthy **(Nyathi mangima ne ber)..HealthSP**….1 Smooth skin, no rashes **(Nyathi ma dende yom, onge gwonyo gwonyo)**…**SkinSP**…… ……1 Improve immunity, prevent illness **(Geng'o/kedo gi tuoche)**……**ImmunSP**……1 Other (**Mamoko)**…**OtherSP**……..1 Don't know **(Akia)**…**DKSP**……1
M17. **OFUKU ACHIEL MAR SPRINKLES EN PESA ADI E GWENG'U KA?** HOW MUCH DOES A SACHET OF SPRINKLES COST IN YOUR COMMUNITY? **SPCost** *(Don't read. Mark only one)*	2 ksh per sachet……..1 5 ksh per sachet……..…2 1.5 ksh per sachet…….....3 1 ksh per sachet…….....4 Other **(Mamoko)**……..…..88 Don't know **(Akia)**……99
M18. **BENDE IPARO NI NG'ENY JI NIGI NYALO MAR NG'IEW SPRINKLES E GWENG'U KA?** DO YOU THINK MOST PEOPLE CAN AFFORD TO BUY SPRINKLES IN YOUR COMMUNITY? **AffordSP**	Yes, it's affordable …… ……..…1 No, not affordable to all …… ….2 It should be free …….. …… ……..3 Other **(Mamoko)**……..88 Don't know **(Akia)**……99

(Don't read. Mark only one)	
M19. **PAKET ACHIEL MAR 'SPRINKLES' IPARO NI ONEGO OBED PESA ADI?** How much do you think one packet of Sprinkles should cost? **ThinkSpCost**	_____ **KSh**
M20. **KAPO NI PAKET ACHIEL MAR 'SPRINKLES' EN SILING' 5 INYALO THORO NG'IEWE BANG' NDALO ADI?** IF THE PRICE OF SPRINKLES IS 5 KSH PER SACHET, HOW OFTEN WOULD YOU BUY THEM? **FreqBuySP** *(Don't read. Mark only one)*	One a day……1 Several times a week……..2 One a week……3 Twice a month……4 One a month……5 A few times a year……6 Never……7 Other……88 Don't know **(Akia)**……99
M21. **IPARO NADE KA PAKET ACHIEL EN SILIN'G ABICH TO IDWARO MIYO NYATHINI DICHIEL KATA DIRIYO E JUMA?** WHAT DO YOU THINK ABOUT THE PRICE OF 1 SACHET FOR 5 KSH IF YOU ONLY NEED TO GIVE IT TO YOUR CHILD **ONCE OR TWICE** A WEEK? **SPOneTwo** *(Don't read. Mark only one)*	Price is OK……0 Price is too high……1 Price is too low……2 Other **(Mamoko)**……88 Don't know **(Akia)**……99
M22. **BENDE SPRINKLES NWANG'ORE MAYOT E GWENG' KA?** DO YOU THINK SPRINKLES ARE EASILY ACCESSIBLE FOR SALE IN YOUR COMMUNITY? **AccessSP**	No **(Ooyo)** ……0 Yes **(Eee)** ……1 Other **(Mamoko)**……88 Don't know **(Akia)**……99

(Don't read. Mark only one)	
M23. **DIHER NG'IEWO 'SPRINKLES KA NYE?** Where would you like to buy sprinkles? *(Don't read. Mark only one)*	SWAP Vendor ……..…..1 Community health worker/promoter…….……2 **Jaus gige SWAP/Nyamrerwa** Pharmacist / chemist **Jaus yedhe/ od yath**……3 Health Facility **Kar thieth**……...4 Retail shops **Dukni** ……....5 Chief's baraza **E barasa** …….……...6 SWAP shop **Duka ming'iewe gige SWAP**……...7 Kiosk **(Kiosko)** ……8 Other……..……..88
M24. **BENDE ISEGA USO SPRINKLES?** HAVE YOU EVER SOLD SPRINKLES? **SoldSP** *(Don't read. Mark only one)*	No **(Ooyo)** ……....0 Yes **(Eee)** …… …… ……..1
M25. **ANGO' MA MONO, KATA MOSE MONO JOMOKO MIYO NYITHINDO SPRINKLES E'GWE U KA?** WHAT ARE THE BARRIERS TO GIVING SPRINKLES TO CHILDREN IN THIS	None **(Onge)**…**BarNone**……1 Cost - including lack of credit **(Nengo ne, onge mar hola)**……**BarCost**…….. … ……..…1 Causes loose stool, diarrhea **(Losruok marep kata diep) BarDiarr**……..……1 Causes increased appetite **(Dhok mamit)**…**BarApp**….1 Parents are lazy, forgetful **(Samuoyo kata wichwil mar**

COMMUNITY? *(Don't read, mark all that apply)*	**jonyuol) BarForget**……1 Child not sick and don't need **(Nyathi ok tuo)**……...1 Meant for children with HIV/AIDs **(Mar nyithindo man gi ayaki)**…**BarHIV**……1 Don't know where to buy **(Akia kama anyalo ngiewe).** 1 Other **(Mamoko)**…**BarOther**……...1 Don't know **(Akia) BarDK**……1
M26. **BER KATA RACH MANE MA ISENENO E NYATHINI (NYITHINDI) BANG' TIYO KOD SPRINKLES?** WHAT POSITIVE OR NEGATIVE EFFECTS DID YOU SEE IN YOUR CHILD(REN) AFTER USING SPRINKLES? *(Don't read, mark all that apply)*	None **(Onge)** ……**EffNone**……..1 Appetizer **(Keto dhok mamit)** …**EffApp**…… …..1 Give energy, make active **(Medo teko) EffEnergy**…..1 Make child, family happy **(Keto nyathi kod jo ot mamor) EffHappy** …… …… … . ……1 Make child playful **(Keto nyathi matugo maber/ma njejre)**…**EffPlay**……..1 Grow healthy, make child healthy **(Nyathi man kod ngima maber)** ..**EffHealth** …… ……1 Improved immunity **(Konyo e geng'o/kedo kod tuoche)**…**EffImmun**…….. …. ……..1 Prevent low blood, adds blood **(Medo remo teko)**…… 1 Make child stronger **(Keto nyathi tegno maber)**……1 Causes diarrhea **(Miyo nyathi diep)** …**EffDiarr**……...1 Causes dark stool **(Keto losruok ma rateng'……1** Causes vomiting **(Kelo ng'ok) .EffVomit……1** Prevent diarrhea **(Geng'o diep)**…**EffNoDiarr**……1 Prevent malaria **(Geng'o malaria/midusi).EffNoMal…**1 Other **(Mamoko)**……**EffOther**……..1 Don't know **(Akia)**……**EffDK**……1
M27. BENDE NE IMIYO NYATHINI SPRINKLES MONDO OTHIEDH NE TUO MORO KANE OTUO? DID YOU EVER GIVE YOUR CHILD SPRINKLES TO TREAT AN ILLNESS WHEN S/HE WAS SICK? **SPRTRTSICK**	No (Ooyo) ……..……0 Yes (Eee) …… …… ……..……1 Don't know (Akia)……..…99
M28. **BENDE ISEYUDO ACHIEL KUOM MAGI?** HAVE YOU EVER RECEIVED ANY OF THE FOLLOWING?	Sprinkles calendar **(Kalenda mar sprinkles)**……0 / 1 Sprinkles leaflet/brochure **(Otase mag lando sprinkles)**… …… 0 / 1 Sprinkles cup **(Okombe mag lando sprinkles)** ……0 / 1 Sprinkles sticker**(Otas mibawo ma lando sprinkles)** 0/1

(Read and mark each one yes or no)	Sprinkles T-shirts **(sprinkles t-shirts)**.....0 /1
M29. **BENDE ISEYUDO SPRINKLES MA OCHIW NONO?** HAVE YOU EVER RECEIVED ANY FREE SPRINKLES FROM: *(Read and mark each one yes or no)*	Launch **(Romo makende mane e lande sprinkles)**..0/ 1 Training **(Tiegruok)**……..**FreeTrn**……0 / 1 Vendor **(Jauso)**……**FreeVen**……..……0 / 1 Neighbor/Friend/Relative **(Jirani/osiepni/watni)**…. 0 / 1 NGO, international agency (e.g., UNICEF)…. 0 / 1
M30. **BENDE NE IDHIYE TIEGRUOK KATA ROMO MAKENDE MI LANDE WECHE MAG SPRINKLES?** DID YOU EVER ATTEND ANY SPRINKLES TRAININGS OR LAUNCHES? **AttendSpr** *(Don't read. Mark only one)*	No **(Ooyo)** ……..……0 Yes **(Eee)** …… …… ……1 Other **(Mamoko)**……..88 Don't know **(Akia)**……99
M31. **Ere yo maber ma inyalo puonj godo mine wach mar sprinkles?** What are the best ways to pass on information about Sprinkles to mothers? *(Don't read. Mark all that apply)*	Radio, T.V. (**Nyakalondo, telebisen)**…… 1 Newspaper ……2 My child in school (**Nyathina mani e skul)**……..3 Brochure / Poster (**Jopuonj mag gweng**)……...4 Promotion show **Lendo mag bath ndara**…… 5

	Community meetings/chiefs baraza **Barasa mar gweng'** ……...6 Truck/loudspeaker **Mtoka man gi aujo** ……..7 Wall painting **Goro mar kor ot**……...8 Health facility **Kar thieth** ……..…..…. 9 Neighbor / family / friends **Jirani/watni/osiepeni** ….. 10 Health Officer/Nurse/CHW **Jaelth/sista/jothieth mantiere e gweng'** **…..**……..……..…… 11 SWAP vendors **Jous gige SWAP**……..……..12 Other **Mamoko** ……..…..88 Don't know (**Akia)** ……..……99

SPRINKLE USE

M32. KUOM JUMBE ARIYO MOSEKALO, OFUKU ADI MAG SPRINKLES MA IN KATA ACHIEL KUOM JOODI OSENG'IEWO KATA OSEYUDO NONO? OVER THE LAST 2 WEEKS, HOW MANY SPRINKLES SACHETS HAVE YOU OR ANYONE IN YOUR HOUSEHOLD PURCHASED OR RECEIVED FOR FREE? **NumSachet**	_____ **sachets**
M33. **BENDE JAODNI MORO AMORA OSETIYO GI SPRINKLES?** HAVE ANY HOUSEHOLD MEMBERS EVER USED SPRINKLES? **SPRINKLE** (DON'T READ. MARK ONLY ONE)	No **(Ooyo)** ……..……..…….....0 Yes **(Eee)** …… ……..…… ……..……1 Don't know **(Akia)**……..…99
M34. **NYISA JOODNI MA JO SWECHE 6-59 MOSETIYO GI SPRINKLES?** PLEASE LIST ANY HOUSEHOLD MEMBERS **6-59**	1. _____

MONTHS OF AGE WHO HAVE EVER USED SPRINKLES	2. _____ 3. _____ 4. _____ 5. _____
M35. **BENDE DANG' ANEE OFUKE MAG SPRINKLES MA IN GODO MA IBIRO TIYO GODO E ODI, KA IN JA USO KIK IKWAN MA IPARO NI IBIRO USO?** Can I see any sprinkles sachets you have available for your household use, do not include any sprinkles you intend to sell if you are a vendor. **SPObs**	Unopened Sprinkles Sachets available……. 1 Unopened Sprinkles Sachets not available..... ..... 2 Opened Sprinkles Sachets available…….. ..…. 3 Refused …….. 99

HHID ☐☐☐☐☐☐☐☐

Child Number ☐☐

CHILD QUESTIONNAIRE (6 MONTHS TO 3 YEARS)

CHILD DEMOGRAPHICS

If the eligible primary caretaker is not present, schedule another visit to the household

C1. NYING NYATHI WHAT IS THE NAME OF THE CHILD?	-----------
C2. NYATHINI ONYUOL KARANG'O? WHAT IS THE CHILD'S DATE OF BIRTH? **CDOB** ***IF DON'T KNOW THE DAY OR MONTH, ENTER 01,01***	☐☐ **Day** ☐☐ **Month** ☐☐☐☐ **Year**
C3. WRITE THE SOURCE OF BIRTH DATE SOURCEDOB	Clinic book (Kad mar klinik) .....0 Baptismal card (Kad mar batiso) .....1 Birth certificate (Barup nyuol) .....2 Recall (Paro gi wich) .....3 Other (Mamok) 88
C4. EN WUOYI KOSO NYAKO	

SEX OF THE CHILD CSEX	Boy (Wuoyi) .....1 Girl (Nyako) 2
C5. NYATHINI EN ANG'ONI? WHAT IS YOUR RELATIONSHIP TO THE CHILD? **CHILDRELN**	Biological Mother **Mingi monyuole** .....1 Female caretaker **Mama marite** .....2 Adoptive mother **Mama mokawe** .....3 Father **Babagi** .....4 Other.....88 Don't know 99
CHILD – Micronutrient Module	
C6. BENDE NYATHINI OSEYUDI GI NOK MAR REMO EDENDE? HAS YOUR CHILD EVER BEEN DIAGNOSED WITH ANAEMIA? ANEMIA	No (Ooyo).....0 Yes (Eee).....1 Don't know (Akia) 99
C7. BENDE SANI OMUONYO/OMADHO YIEN MAG NOK MAR REMO E DE? IS THE CHILD CURRENTLY TAKING IRON SUPPLEMENTS (E.G., RB TONE, NOT SPRINKLES)? CHILDIRON	No (Ooyo) .....0 Yes (E ee) .....1 Don't know (Akia).....99
C8. NOTIYO GI YIEND MEDO REMO DIDI E JUMA MOKALO? HOW MANY TIMES DID YOUR CHILD TAKE IRON SUPPLEMENTS (E.G., RB TONE, NOT SPRINKLES) IN THE LAST WEEK? TimesIron	☐☐ Number of times *(IF 'DON'T KNOW', ENTER 99)*
C9. ANG'O MOMIYO NYATHINI OK TI GI YIEN MAMEDO REMO SANI? WHY IS YOUR CHILD NOT TAKING IRON SUPPLEMENTS (E.G., RB TONE, NOT SPRINKLES) CURRENTLY? NOIRON *(DON'T READ. MARK ONLY ONE)*	Child does not need it; he is healthy **(Onge tiende, nyathi ngimane ber**).....1 Terminated treatment **(Osetieko thieth)** .....2 Do not have money to buy iron **(Aonge pesa mar ng'iewo)** .....3 Child had an adverse reaction to iron (**Okelo tabu e dend nyathi)** .....4 Child has not been able to see medical provider **(Nyathi pok oneno laktar**) .....5 Do not have access to iron (**Onge kama iyudo e yedhe go)** .....6 Other, specify (**Mamoko)** .....88 Don't know (**Akia**) 99

IF NO, GO TO C9

CHILD – Breastfeeding Module		
C10. **BENDE (NYING) OSEGA DHOTH?** HAS THE CHILD EVER BEEN BREASTFED OR BEEN FED BREAST MILK? **EVERBREAST**	No **(Ooyo)** .....0 Yes **(Eee)** .....1 Refused (**Notamore)** .....77 Don't know **(Akia)** 99	**IF NO OR DK, GO TO C13**
C11. **KACHAKRE NYORO SECHE MACHALO GI MAGI BENDE (NYING) OSEDHOTH?** SINCE YESTERDAY, A TIME LIKE THIS, HAS THE CHILD BREASTFED? **BREASTYEST**	No (**Ooyo)** .....0 Yes (**Eee)** 1	
C12. **NYATHINI NOWEYO DHOTH KAJA HIGNI ADI?** AT WHAT AGE DID YOU STOP BREASTFEEDING THE CHILD? **StopBrMon**	☐☐ Months *If don't know then '99'* *If still breastfeeding then '66'*	
C13. **CHAKRE NYORO SAA MACHALO KAMA, BENDE NYATHINI NOSE MADHO CHAE?** SINCE YESTERDAY, AT A TIME LIKE THIS, DID THE CHILD DRINK ANY TEA? **TEAYEST**	No **(Ooyo)**.....0 Yes (**Eee**).....1 Don't know **(Akia)**.....99	
C14. **CHAKRE NYORO SAA MACHALO KAMA, BENDE NYATHINI OSECHAMO CHILO, BURU, LOWO KATA ODOA?** SINCE YESTERDAY, AT A TIME LIKE THIS, HAS THE CHILD EATEN DIRT, EARTH, OR ODOA? **EATEARTH**	No **(Ooyo)**.....0 Yes (**Eee**).....1 Don't know **(Akia)** 99	**IF NO OR DK, GO TO C16**
C15. **KUOM NDALO ABIRIO MOKALO GIN NDALO ADI____ MANE NYATHINI CHAMO CHILO,BURU,LOWO KATA ODOA?** OVER THE LAST WEEK (7 DAYS), ON HOW MANY DAYS DID THE CHILD EAT DIRT, EARTH, OR ODOA? **DAYSEARTH**	☐☐ days (*If don't know then '99'*)	

CHILD – Malaria & general health Read: "Koro adwaro penji weche kaluwore gi ngima mar nyathini" **Now I'm going to ask you a few questions about the health of your child"**	
C16. **BENDE NYATHINI OSEBEDO KA DIEWO KUOM NDALO ACHIEL MOKALO?** HAS THIS CHILD HAD DIARRHEA IN THE LAST 24 HOURS? **(≥3 LOOSE OR WATERY STOOLS IN A 24 HOUR PERIOD)**	No **(Ooyo)**.....0 Yes (**Eee**).....1 Don't know **(Akia)**..... 99
C17. **BENDE OSEBEDO GI TUO MAR KOR MATHUNG' KATA AHONDA KUOM**	No **(Ooyo)**.....0 Yes (**Eee**).....1

NDALO ACHIEL MOKADHO? RESP24H HAS THIS CHILD **HAD** RESPIRATORY ILLNESS IN THE LAST 24 HOURS? **(COUGH OR BREATHING PROBLEMS)**	Don't know **(Akia)** 99	
C18. **BENDE OSEBEDO GI DEL MAORE KUOM NDALO ACHIEL MOKADHO?** HAS THIS CHILD HAD A FEVER IN THE LAST 24 HOURS? **FEVER24H**	No **(Ooyo)**.....0 Yes (**Eee**).....1 Don't know **(Akia)..** 99	
C19. **BENDE OSEBEDO GI MALARIA EJUMBE ARIYO MOKALO? MAL2WKS** HAS THIS CHILD HAD MALARIA DURING THE LAST 2 WEEKS?	No **(Ooyo)**.....0 Yes (**Eee**).....1 Don't know **(Akia)** 99	
C20. **BENDE NYATHINI OSENINDO E HOSPITAL KUOM JUMBE ARIYO MOKADHO? HOSP2WKS** HAS THIS CHILD BEEN HOSPITALIZED IN THE LAST 2 WEEKS (14 DAYS)?	No (**Ooyo**).....0 Yes **Eee)**.....1 Don't know **(Akia)** 99	**IF NO OR DK, GO TO C22**
C21**. NE EN GI CHANDRUOK MANE?** WHAT WAS THE HEALTH PROBLEM? **HOSPHPROB**	Diarrhea (**Diep**) ..... 1 Respiratory infection (**Kor mathung'**) ..... 2 Malaria **(Mhidusi)** ..... 3 Other **(Mamoko)** .....88 Don't know **(Akia)..** 99	
C22. **BENDE NYATHINI NONINDO E BUO NET NYORO GOTIENO?** DID (NAME) SLEEP UNDER A MOSQUITO NET LAST NIGHT? **CHLDSLPITN**	No (**Ooyo**).....0 Yes **Eee)**.....1 Don't know **(Akia)** 99	
SPRINKLES USE MODULE		
C23. **BENDE NGANI OSETIYO GA GI SPRINKLES?** HAS (NAME) EVER USED SPRINKLES? **SPRKUSEEVER**	No (**ooyo**).....0 Yes (**Eee**).....1 Don't know (**Akia**).....99	
C24. **CHAKRE ODIECHIENG' MANYORO NYAKA SANI (KAWUONO) BENDE ____ OSETIYO GI SPRINKLES?** SINCE YESTERDAY UNTIL NOW—TODAY, HAS THIS MEMBER USED SPRINKLES? SprkUseYest	No **(Ooyo)** .....0 Yes **(Eee)** ....1 Don't know **(Akia)**.....99	
C25. **KUOM NDALO ABIRIYO MOSEKALO KOCHAKORE KAWUONO, NG'ANI OSETIYO GI SPRINKLES ADI?** STARTING WITH TODAY, OVER THE LAST 7 DAYS HOW MANY SPRINKLES SACHETS DID <CHILD'S NAME> CONSUME? **SPRKUSE7DAYS**	☐☐ sachets	

C26. CHAKRE KAWUONO, KIDOK CHIEN NDALO ABIRIYO MOSEKALO, NDALO ADI MA (NG'ANI) OSETIYO GI SPRINKLES? Starting with today, over the last 7 days on how many days has <child's name> used Sprinkles? SprkDays7Days	☐☐ Days

*Enumerator: Is there another **SELECTED** child 6-35 months that lives in this household?*

*If **Yes,** Fill out another CHILD Questionnaire If **No**, end of survey*

---------- That is the last question. Thank you for answering our questions. ----------

Appendix 2: Expanded Literature Review

I. The search for a gold-standard of iron indicators and the effect of inflammation

1. Grant et al. in AJCN (2012) – Comparision of iron indicators

From March to May 2009, Grant et al. conducted a community-based cluster design survey of 680 children aged 6-35 months across 60 randomly selected villages in western Kenya in which they also measured indicators of iron status (ferritin, TfR, and ZP; and the calculated TfR/ferritin index) from capillary blood samples. Their research question was to determine if a single iron biomarker could be identified as adequate to replace the multiple-criteria model for screening for iron deficiency at the population level, particularly in the low resource setting. The multiple-criteria model identifies iron deficiency if two or more iron indicators indicate deficiency. The authors used the kappa statistic to measure the extent of agreement between each iron indicator (or the index) and the multiple-criteria model. The results presented in this study were unadjusted for inflammation, however the authors mentioned that their conclusions were unchanged when they did adjust for inflammation. They found that the kappa statistic was highest for TfR (0.88, 95% CI: 0.84-0.92) and lowest for ferritin (0.35, 95% CI 0.30-0.40). A kappa statistic of >0.75 indicated excellent agreement.

The authors also calculated sensitivity and specificity for each measure of iron status in identifying iron deficiency (as defined by the multiple-criteria model) and found that sensitivity was greatest for ZP (0.99 ± 0.01) and TfR (0.95 ± 0.01), and lowest for ferritin (0.43 ± 0.02), while specificity was greatest for TfR (0.94 ± 0.01) and lowest for ZP (0.44 ± 0.03). Because TfR had a high sensitivity and specificity, the authors concluded that TfR least misclassified children as iron deficienct.

2. Grant et al. in Journal of Nutrition (2012) – Inflammation correction factors

There are several possible approaches to account for inflammation in measuring iron status: 1) upward adjustment of cutoff values, 2) exclusion of individuals with elevated inflammatory markers, 3) use of a calculated correction factor for inflammation based on the classification of four levels of inflammation, and 4) adjustment in regression analysis. This paper explores the use of inflammatory correction factors among the same population as described above in Grant et al. in AJCN (2012). CRP is an early sign of inflammation and becomes elevated in the first 10 hours of acute inflammation, reaches its peak at about 48 hours, and normalizes within one week. AGP, on the other hand, begins to rise about 24 hours after the onset of inflammation and has a slower decline. Therefore, four levels of inflammation were defined: reference (normal CRP and AGP), incubation (elevated CRP, normal AGP), early convalescence (elevated CRP and AGP), and late convalescence (normal CRP and elevated AGP). Geometric means of each iron indicator (ferritin, ZP, TfR, TfR/ferritin index) were all statistically significantly different by inflammation status. For each inflammation group, correction factors were calculated as the ratio of the geometric means of iron indicators compared to the reference group. Iron indicators were then "corrected" by multiplying the true value by the group-specific correction factor. The use of corrected iron indicators instead of the uncorrected values, in particular for ferritin, resulted in an increased

estimated prevalence of iron deficiency among this population of children (from an estimated prevalence of 27% when uncorrected to 41% when corrected). The correlation, measured by Spearman correlation coefficient, was also strengthened when corrected ferritin was used (from 0.10 to 0.20).

3. Kung'u et al. in Journal of Nutrition (2009) – Adjusting for inflammation with iron indicators

Another approach to accounting for inflammation status in measurements of iron deficiency is model adjustment for inflammation. In this paper, Kung'u et al. used a subsample of children aged 6-23 months previously recruited for a clinical trial in Zanzibar that consisted of 230 age-matched triplets (690 children total) selected based upon infection status at a ratio of 2:1 (2 helminth-infected for 1 unaffected child). This was a cross-sectional analysis of the archived biochemical indices of these 690 children. As would be expected, the authors found a moderate positive correlation between ferritin and the acute phase reactants AGP and CRP and a weak positive correlation with ZP and TfR. When linear regression was used to model the relationship between hemoglobin and ferritin, simple linear regression showed a negative slope. However, with the addition of recent fever, CRP, and AGP to the model, the relationship became reversed to a positive relationship, as would be expected. This finding reaffirms the importance of accounting for inflammation when using ferritin as an iron indicator. However, when ZP and TfR were modeled by simple linear regression with Hb, the direction of the relationship between Hb and ZP or TfR was not affected by CRP or AGP. The magnitude of the relationship between Hb and ZP and TfR, respectively, was slightly elevated when adjusted for inflammation. The authors concluded the ZP and TfR were minimally influenced by CRP and AGP, while ferritin was greatly affected. They also concluded that using recent fevers alone as a proxy for inflammation status was not sufficient; at least one biomarker for inflammation, CRP or AGP, should be used to adjust the iron indicators for the influence of inflammation.

II. Inherited blood disorders and anemia:

1. Suchdev et al. in Maternal and Child Nutrition (2012) – The prevalence of hemoglobinopathies in western Kenya

This was the same study population as used in our analysis. A total of 858 children were included in the final analysis. This study was important because it reports the prevalence of inherited blood disorders in our population of children aged 6-35 months in western Kenya, which was previously thought to be high but was not quantified. The authors reported that more than two-thirds of the study population had at least one measured blood disorder. More specifically, 17.1% had sickle cell trait and 1.6% had sickle cell disease; 38.5% were heterozygous for alpha-thalasemmia and 9.6% were homozygotes; 20.4% of children had haptoglobin 2-2 genotype; and finally, 6.8% of children had G6PD deficiency.

There was not a statistically significant difference in the distribution of malaria by the measured blood disorder, except that boys with G6PD definicency were less likely to have clinical malaria compared to

boys with a normal G6PD genotype (this trend did not hold among girls). Alpha-thalassemia was associated with an increased prevalence of anemia (definced as Hb < 110 g/L), both before and after excluding children with other known causes of anemia (malaraia parasitemia, inflammation (CRP >5 mg/L), iron deficiency (SF<12 ug/L) or vitamin A deficiency (RBP <0.7 ug/l)). Additionally, the hemoglobin concentration was 4-6 g/L lower among alpha-thalassemics compared to normal children.

III. Inherited blood disorders and iron biomarkers:

1. George et al. in Journal of Nutrition (2012) – Linear regression analysis of hemoglobinopathies and iron indicators and hemoglobin

George et al. conducted a 2-stage cluster-designed, community-based, cross-sectional survey of 2168 children aged 6-59 months from 3 rural provinces in Cambodia and the urban Municipality of Phnom Penh. The questionnaire covered information on the child's health status, household characteristics, and anthropometric measurements. Blood samples were collected for measurement of hemoglobin level, analysis of hemoglobin type (including 4 variants: Hb E trait and disease, alpha-thalassemia trait and disease), measurement of iron status (ferritin, TfR) and vitamin A status (RBP), and quantification of inflammation (AGP, CRP).

Multiple linear regression analysis was used to identify the independent predictors of hemoglobin, ferritin, and TfR. Focusing on inherited blood disorders, the 4 abnormal Hb variants were all significantly associated with lower hemoglobin concentrations (beta-coefficients: Hb E trait, -3.72; alpha-thalasemmia trait, -2.45; Hb E trait with alpha-thalassemia trait, -4.36; and Hb E disease, -11.8); 3 of the abnormal Hb variants were significantly associated with higher log ferritin (beta-coefficients: Hb E trait, 0.13; Hb E trait with alpha-thalassemia trait, 0.13; Hb E disease, 0.49), and with higher log TfR (Hb E trait, 0.03, alpha-thalassemia trait, 0.04; and Hb E disease, 0.26). These models were adjusted for inflammation. Of note, the R-squared values in this study were higher than those in our analysis, likely due to the inclusion of the other iron indicators in the models. All of the hemoglobin variants were statistically significantly associated with an increased risk of anemia, compared to children with a HbAA genotype. This study highlights the complexity of using ferritin and TfR to assess iron status in a population with a high prevlance of hemoglobinopathies.

2. Thurlow et al. in American Journal of Clinical Nutrition (2005)

This group conducted a cross-sectional study of 567 children aged 6-13 years from northeast Thailand in June and July 2002 in which they collected blood samples to measure hemoglobin concentration, mean corpuscular volume, hemoglobin type (HbAA, HbAE, HbEE), and iron status indicators (ferritin, TfR), as well as retinol, vitamin B12, and erythrocyte folate concentrations. Of note, children with a CRP>=10 mg/L were excluded from analysis (n=12). The authors found that HbEE children had significantly higher ferritin and TfR concentrations compared to the children with HbAA or HbAE. ANOVA showed that in

children with HbAA and HbAE (excluding HbEE), age, hemoglobin type, and serum retinol (but not ferritin) were important predictors of hemoglobin concentration. When looking at the population of children with HbAA and HbAE stratified by anemia (anemia defined as hemoglobin concentration of <115 g/L for children aged 6-11 y and <120 g/L for children aged >= 12 y), adjusting for age, anemic children has significantly lower hematocrit, MCV, and serum retinol levels than children who were not anemic; however, mean ferritin and TfR was not significantly different between these two groups. This finding indicates that little of the anemia in this population was related to storage iron depletion. Researchers also found that in the HbAA and HbAE population, anemic children had lower serum retinol level compared to their non-anemic counterparts. Stratifying by vitamin A status (deficiency defined as serum retinol level <0.7 umol/L) and adjusting for age, they found that serum ferritin was significantly higher among children with vitamin A deficiency while there was no difference in mean TfR between the two groups. The authors postulated that vitamin A deficiency may result in a reduction in hematopoiesis due to decreased mobilization of iron from spleen or liver stores into circulation, resulting in higher serum ferritin concentrations despite higher rates of anemia in the setting of vitamin A deficiency.

3. Rees et al. in British Journal of Haematology (1998)

As part of a community-based survey of malaria on the South-Western Pacific island of Santo in Vanuatu, the researchers recruited 181 age- and sex-matched children by alpha thalassemia genotype. They measured the concentrations of soluble transferrin receptor and ferritin. Multiple linear regression analysis was used to determine the significant predictors of log TfR. They found that the homozygous alpha thalassemia genotype, the heterozygous alpha thalassemia genotype, and log ferritin were all significant predictors of log TfR, and together accounted for 13% of the variability in log TfR. The authors postulate that the association between alpha thalassemia deletions and TfR is explained by alpha thalassemia causing ineffective erythropoiesis, and therefore an increase in TfR.

IV. Pemba controversy:

1. Sazawal et al. in The Lancet (2006): Pemba Trial

Sazawal et al. recruited 32,155 children between the ages of 1-35 months to a four-arm cluster-randomized double-masked supplementation efficacy study beginning in January 2002 in Pemba, Zanzibar. The four arms were supplementation with a) iron and folic acid, b) iron, folic acid, and zinc, c) zinc alone, d) placebo. The outcomes of interest were all-cause mortality and hospital admission. The study was discontinued in August 2003 because the incidence of serious adverse events (including death or hospital admission) in Arms A and B combined was 12% higher than that of the placebo group (p=0.02).

2. Sazawal et al. in The Lancet (2006): Pemba Trial Substudy

In this substudy of 2,413 children (or about 800 per arm), results of iron and folic acid supplementations were stratified by measures of iron status assessed by zinc protoporphyrin. Iron deficiency was defined as ZP>80umol/mol and 75% of the substudy population was identified as iron deficient according to this cut-off.

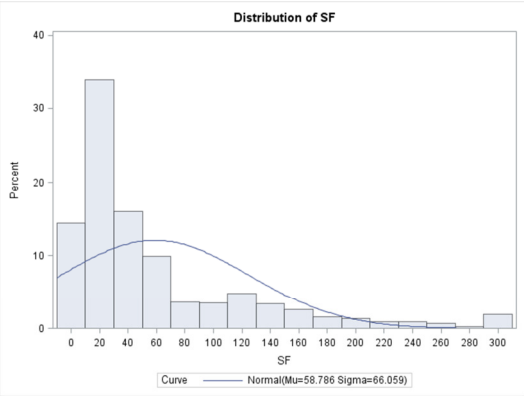
Overall, treatment Arms A and B trended towards fewer hospital admission and fewer deaths compared to the placebo arm, however these results were not statistically significant. They found that children in Arms A and B who were iron deficient experienced a reduction in risk of all-cause mortality and hospital admission, compared to the placebo arm (RR 0.62, 95% CI 0.41-0.93, p=0.02). However, children who were not iron deficient who were in Arms A and B had a non-statistically significant increased risk compared to the placebo arm (RR 1.63, 95% Ci 0.72-3.66, p=0.24).

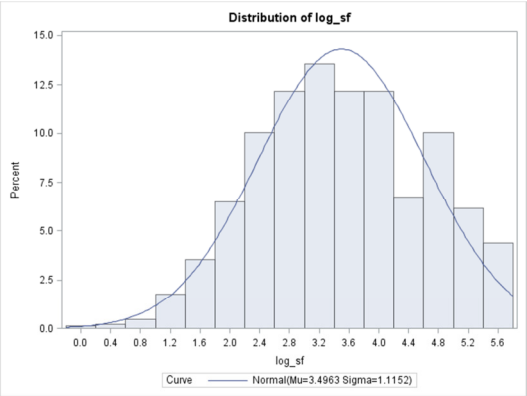
3. Tielsch et al. in The Lancet (2006): Nepal Trial

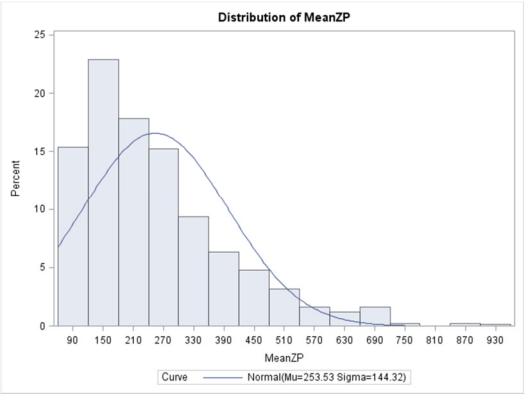
The Nepal Trial was very similar to the Pemba Trial in design and outcomes of interest, however it was conducted in an area of low malaria prevalence. This was a community-based, cluster-randomized, double-masked, placebo-controlled trial conducted among a study population of 25,490 children aged 1-36 months in the southern plains of Nepal that began in October 2001 and was stopped in November 2003 due to no evidence of a beneficial effect and low power. The four study arms were the same as those in Pemba. The primary outcome was all-cause mortality, and secondary outcomes were cause-specific mortality, as well as the incidence and severity of diarrhea, dysentery, and acute respiratory illnesses. Iron deficiency was defined as a serum ferritin level <12 ug/L. They found that there was no difference in mortality (primary outcome) or rates of diarrhea, dysentery, or respiratory illnesses between Arms A and B compared to the placebo arm.

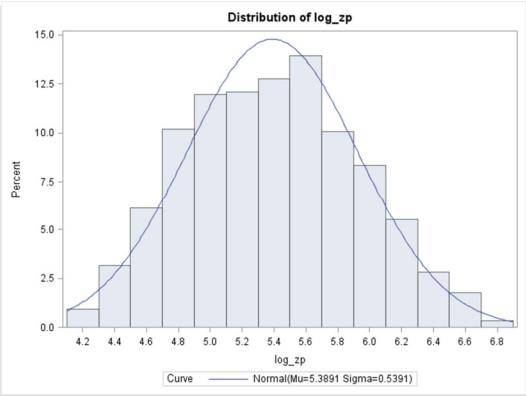
Appendix 3: Distribution plots for iron indicators: Pre- and post-log transformation.

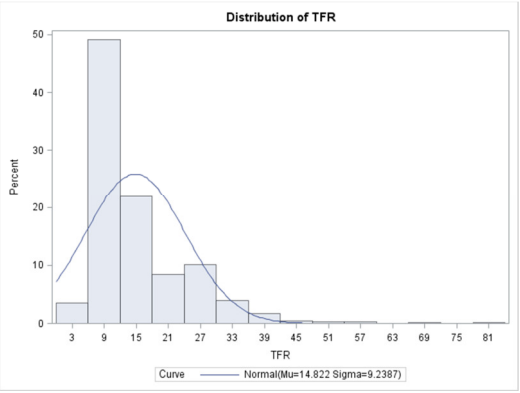
Untransformed: Ln Transformed:

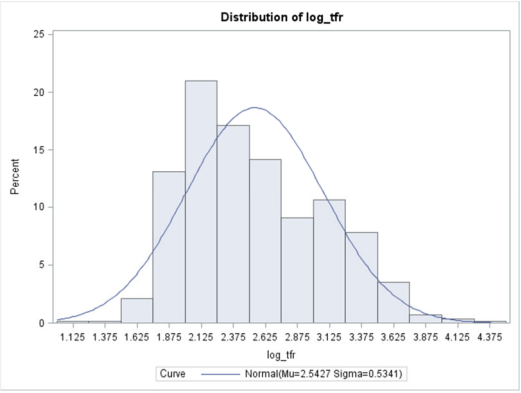


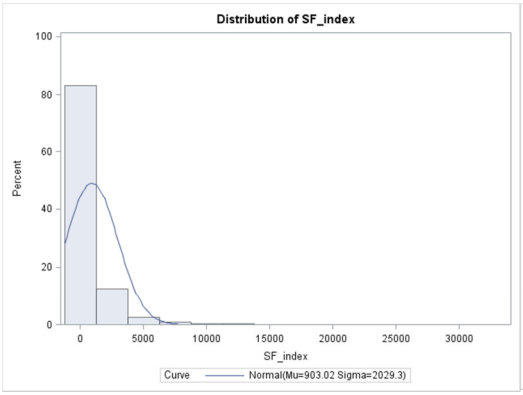


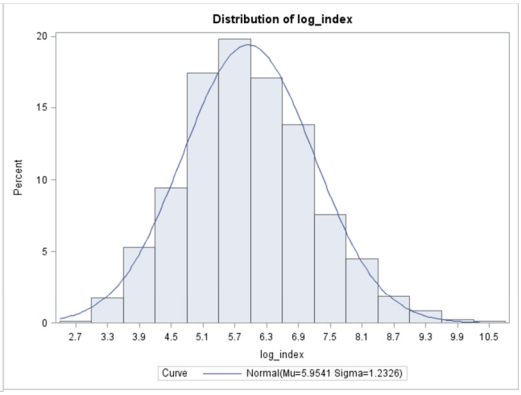


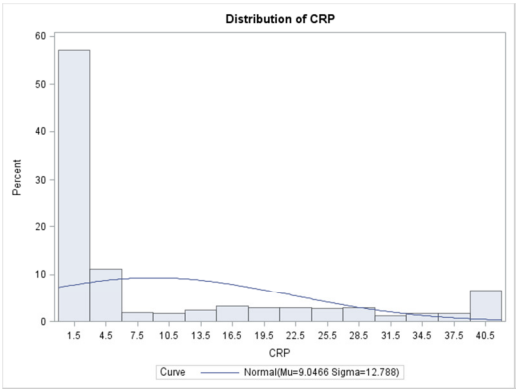


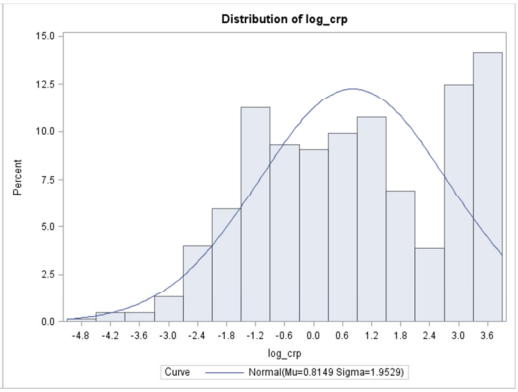


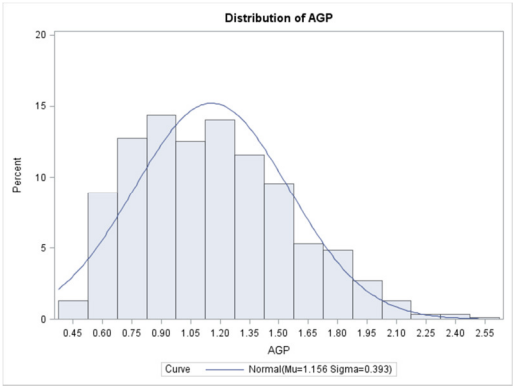


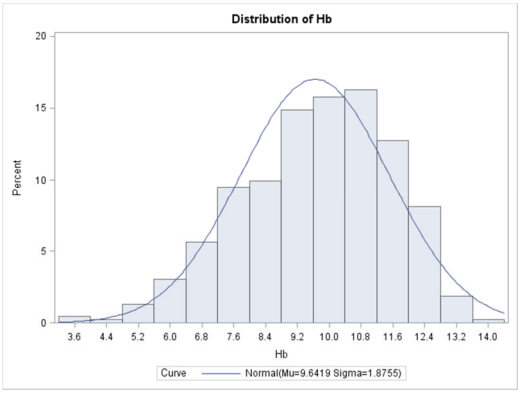

Distribution of SF_index
Percent
100
80
60
40
20
0
0
5000
10000
15000
20000
25000
30000
SF_index
Curve
Normal(Mu=903.02 Sigma=2029.3)

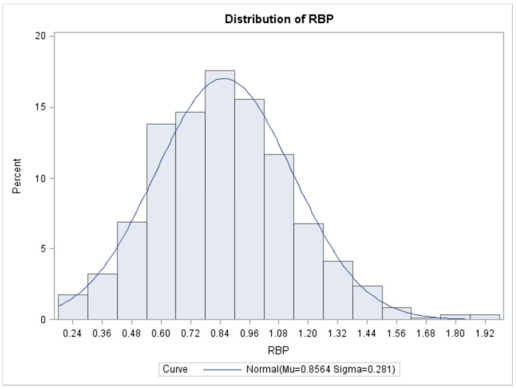

Distribution of log_index
Percent
20
15
10
5
0
2.7
3.3
3.9
4.5
5.1
5.7
6.3
6.9
7.5
8.1
8.7
9.3
9.9
10.5
log_index
Curve
Normal(Mu=5.9541 Sigma=1.2326)


Distribution of CRP
Percent
60
50
40
30
20
10
0
1.5
4.5
7.5
10.5
13.5
16.5
19.5
22.5
25.5
28.5
31.5
34.5
37.5
40.5
CRP
Curve
Normal(Mu=9.0466 Sigma=12.788)


Distribution of log_crp
Percent
15.0
12.5
10.0
7.5
5.0
2.5
0.0
-4.8
-4.2
-3.6
-3.0
-2.4
-1.8
-1.2
-0.6
0.0
0.6
1.2
1.8
2.4
3.0
3.6
log_crp
Curve
Normal(Mu=0.8149 Sigma=1.9529)


Distribution of AGP
Percent
20
15
10
5
0
0.45
0.60
0.75
0.90
1.05
1.20
1.35
1.50
1.65
1.80
1.95
2.10
2.25
2.40
2.55
AGP
Curve
Normal(Mu=1.156 Sigma=0.393)


Distribution of Hb
Percent
20
15
10
5
0
3.6 4.4 5.2 6.0 6.8 7.6 8.4 9.2 10.0 10.8 11.6 12.4 13.2 14.0
Hb
Curve
Normal(Mu=9.6419 Sigma=1.8755)


Distribution of RBP
Percent
20
15
10
5
0
0.24 0.36 0.48 0.60 0.72 0.84 0.96 1.08 1.20 1.32 1.44 1.56 1.68 1.80 1.92
RBP
Curve
Normal(Mu=0.8564 Sigma=0.281)

Appendix 4a: Assessing confounding and interaction for the model of ZP and sickle cell

Model	Column1	Independent variables	Beta 1	Independent variable p-value	INTERACTION?	Crude B1 (from SLR)	p-value	Adjusted B1	p-value2	10% Rule (output <0.90?)	CONFOUNDING?
ZP and Sickle Cell	Age	Sickle cell	0.0135929	0.92		-0.0434047	0.36000	-0.04214		0.97086721	No
		Age	-0.0022746	0.34							
		Age*Sickle	-0.0025618	0.64	No						
	Sex	Sickle cell (x1)	0.01518711	0.82		-0.0434047	0.36	-0.04203		0.968330619	No
		Sex (ref = female)	0.08587573	0.36							
		Sex*sickle	0.05414759	0.60	No						
	Inflammation stage	Sickle cell	-0.0008276	0.99		-0.0434047	0.36	-0.036047		0.830488403	**Yes**
		Inflammation	0.2921645	**<0.0001**					Inflammation:		
		Sickle*Inflammation	-0.0588205	0.63	No						
	RBP	Sickle cell	-0.1506475	0.33		-0.0434047	0.36	-0.041631	Sickle:	0.959133458	No
		RBP	-0.4208304	**0.04**					RBP:		
		RBP*Sickle cell	0.127598	0.42	No						
	Stunting	Sickle	-0.0194244	0.72		-0.0434047	0.36	-0.051402	Sickle:	1.18425424	**Yes**
		Stunting	0.1497961	**0.0017**					Stunting:		
		Stunting*Sickle	-0.0963123	0.25	No						
	Wasting	Wasting	0.1435884	0.16		-0.0434047	0.36	-0.040805		0.94011478	No
		Sickle cell	-0.0381458	0.43							
		Wasting*Sickle cell	-0.0995445	0.63	No						
	Underweight	Underweight	0.1102985	0.08		-0.0434047	0.36	-0.046413		1.069308162	No
		Sickle cell	-0.0622301	0.19							
		Underweight*Sickle cell	0.1148724	0.41	No						
	SES	SES	-0.0152672	0.73		-0.0434047	0.36	-0.041784		0.962658422	No
		Sickle cell	-0.0022579	0.97							
		SES*Sickle	-0.1037399	0.27	No						
	Maternal Education	Maternal Education	-0.0351659	0.41		-0.0434047		-0.034023	Sickle: 0.48	0.78386442	**Yes**
		Sickle cell	-0.0164257	0.81					Edu: 0.26		
		Maternal Edu*Sickle	-0.0334827	0.71	No						
	Recent tea	Recent tea	-0.0213064	0.66		-0.0434047		-0.038975	Sickle:	0.897941928	**Yes**
		Sickle cell	-0.2035475	0.06					Tea:		
		Recent tea*sickle	0.1997047	0.09	No - borderline						
	Recent Sprinkles	Recent Sprinkles	0.0357259	0.59		-0.0434047	0.36	-0.041966		0.966849212	No
		Sickle cell	-0.0124138	0.81							
		Recent Sprinkles*sickle	-0.2391695	**0.04**	**Yes**						
	Malaria parasitemia	Malaria parasitemia	0.3858808	**<0.0001**		-0.0434047	0.36	-0.046572		1.072975968	No
		Sickle cell	0.054651	0.42							
		Malaria*sickle	-0.3011018	**0.0031**	**Yes**						
	Recent fever	Recent fever	0.1886236	**<0.0001**		-0.0434047	0.36	-0.041784		0.962656118	No
		Sickle cell	0.0064162	0.92							
		Recent fever*sickle	-0.1147798	0.27	No						
	Thalassemia	Thal	0.0386628	0.42		-0.0434047	0.36	-0.038937	Sickle: 0.4	0.897073358	**Yes**
		sickle cell	-0.014701	0.82					Thal: 0.42		
		Thal*Sickle	-0.0481693	0.64	no						
	G6PD	G6PD	-0.1026853	0.1492		-0.0434047	0.36	-0.052147	Sickle: 0.28	1.201420583	**Yes**
		sickle cell	-0.0514005	0.2974					G6PD: 0.15		
		G6PD*sickle	-0.0100112	0.9521	no						

Appendix 4b: Assessing confounding and interaction for the model of TfR and sickle cell

Model	Column1	Independent variables	Beta 1	Independent variable p-value	INTERACTION	Crude B1 (from SLR)	p-value	Adjusted B1	p-value2	10% Rule	CONFOUNDING?
TfR and Sickle Cell	Age	Sickle cell	0.1092822	0.42		0.06872484	0.23	0.0692816	Sickle: 0.22	1.008101292	No
		Age	-0.0011321	0.58					Age: 0.49		
		Age*Sickle	-0.001844	0.73	No						
	Sex	Sickle cell (x1)	-0.056453	0.45		0.06872484	0.23	0.0691546	Sickle: 0.22	1.006253489	No
		Sex (ref = female)	0.1684795	0.1					Sex: <0.0001		
		Sex*sickle	-0.0253756	0.82	No						
	Inflammation	Sickle cell	0.0861617	0.26		0.06872484	0.23	0.0754354	Sickle: 0.16	1.09764388	**Yes**
		Inflammation (vs reference)	0.3045487	**<0.0001**					Inflammation: <0.0001		
		Sickle*inflammation	-0.0177148	0.8707	No						
	RBP	Sickle cell	-0.0763215	0.61		0.06872484	0.23	0.0678842	Sickle: 0.22	0.987768033	No
		RBP	-0.2685992	**0.001**					RBP: 0.003		
		RBP*Sickle cell	0.1687671	0.29	No						
	Stunting	Sickle	0.1051091	0.09		0.06872484	0.23	0.0573414	Sickle: 0.32	0.834361928	**Yes**
		Stunting	0.1975883	**<0.0001**					Stunting: <0.0001		
		Stunting*Sickle	-0.1425008	0.18	No						
	Wasting	Wasting	0.1494989	0.08		0.06872484	0.23	0.0703953	Sickle: 0.22	1.024306641	No
		Sickle cell	0.0757501	0.18					Wasting: 0.12		
		Wasting*Sickle cell	-0.1981705	0.38	No						
	Underweight	Underweight	0.16407219	**0.01**		0.06872484	0.23	0.0640854	Sickle: 0.25	0.932492531	No
		Sickle cell	0.06040865	0.26					Under: 0.01		
		Underweight*Sickle cell	0.02654389	0.86	No						
	SES	SES	-0.0089474	0.82		0.06872484	0.23	0.0708668	Sickle: 0.21	1.031167188	No
		Sickle cell	0.0927563	0.2					SES: 0.60		
		SES*Sickle	-0.0568768	0.58	No						
	Maternal Education	Maternal Education	-0.0592195	0.13		0.06872484	0.23	0.0775505	Sickle: 0.17	1.128420233	**Yes**
		Sickle cell	0.096067	0.21					Edu: 0.07		
		Maternal Edu*Sickle	-0.0352484	0.69	No						
	Recent tea	Recent tea	0.0073737	0.85		0.06872484	0.23	0.0732353	Sickle: 0.19	1.065631	No
		Sickle cell	-0.0535891	0.68					Tea: 0.34		
		Recent tea*sickle	0.154231	0.26	No						
	Recent Sprinkles	Recent Sprinkles	0.0572965	0.35		0.06872484	0.23	0.0723465	Sickle: 0.21	1.052697976	No
		Sickle cell	0.1040064	0.08					Sprinkles: 0.97		
		Recent Sprinkles*sickle	-0.2532357	**0.02**	**Yes**						
	Malaria parasitemia	Malaria parasitemia	0.4062187	**<0.0001**		0.06872484	0.23	0.062041	Sickle: 0.26	0.902744917	Yes (borderline)
		Sickle cell	0.1533549	0.04					Malaria: <0.0001		
		Malaria*sickle	-0.2687627	**0.02**	**Yes**						
	Recent fever	Recent fever	0.1513322	**0.001**		0.06872484	0.23	0.0697312	Sickle: 0.22	1.014642741	No
		Sickle cell	0.0908468	0.21					Fever: 0.0004		
		Recent fever*sickle	-0.0497598	0.69	No						
	Thalassemia	Thal	0.0775389	0.09		0.06872484	0.23	0.0636086	Sickle: 0.26	0.925553992	No
		sickle cell	0.0640188	0.39					Thal: 0.08		
		Thal*Sickle	-0.00082	0.99	no						
	G6PD	G6PD	-0.0297655	0.64		0.06872484	0.23	0.056548	Sickle: 0.32	0.822817485	**Yes**
		sickle cell	0.0625554	0.28					G6PD: 0.46		
		G6PD*sickle	-0.0796213	0.65	no						

Appendix 4c: Assessing confounding and interaction for the model of SF and sickle cell

Model	Column1	Independent variables	Beta 1	Independent variable p-	INTERACTION?	Crude B1 (from SL	p-value	Adjusted B	p-value2	10% Rule	CONFOUNDING?
SF and Sickle Cell	Age	Sickle cell	0.1138467	0.67		-0.0273823	0.23	-0.030991	Sickle: 0.75	1.131785862	**Yes**
		Age	0.0106411	**0.04**					Age: 0.05		
		Age*Sickle	-0.0066771	0.55	No						
	Sex	Sickle cell (x1)	0.0704839	0.63		-0.0273823	0.23	-0.027918	Sickle: 0.77	1.019552777	No
		Sex (ref = female)	-0.1150178	0.55					Sex: 0.03		
		Sex*sickle	-0.0850401	0.69	No						
	Inflammation stage	Sickle cell	0.052482	0.66		-0.0273823	0.23	-0.001804	Sickle: 0.98	0.065881975	**Yes**
		Inflammation	1.1650348	**<0.0001**					Inflammation: <0.0001		
		Sickle*Inflammation	-0.089655	0.6	No						
	RBP	Sickle cell	-0.3889478	0.22		-0.0273823	0.77	-0.0307	Sickle: 0.75	1.121147603	**Yes**
		RBP	-1.014178	**<0.0001**					RBP:<0.0001		
		RBP*Sickle cell	0.4192655	0.22	No						
	Stunting	Sickle	-0.0194906	0.86		-0.0273823	0.77	-0.027741	Sickle: 0.77	1.013085095	No
		Stunting	0.0025567	0.97					Stunting: 0.97		
		Stunting*Sickle	-0.0246115	0.91	No						
	Wasting	Wasting	0.7664418	**0.0004**		-0.0273823	0.77	-0.021044	Sickle: 0.83	0.768540261	**Yes**
		Sickle cell	0.0097	0.92					Wasting: 0.01		
		Wasting*Sickle cell	-1.1378002	**0.04**	**Yes**						
	Underweight	Underweight	0.5282418	**0.001**		-0.0273823	0.77	-0.040621	Sickle: 0.68	1.483480204	**Yes**
		Sickle cell	0.0839779	0.41					Under: 0.02		
		Underweight*Sickle cell	-0.8995285	**0.001**	**Yes**						
	SES	SES	-0.0312868	0.75		-0.0273823	0.77	-0.042077	Sickle: 0.66	1.536649588	**Yes**
		Sickle cell	-0.2614291	0.07					SES: 0.37		
		SES*Sickle	0.5699559	**0.01**	**Yes**						
	Maternal Education	Maternal Education	-0.145545	0.14		-0.0273823	0.77	-0.049792	Sickle: 0.61	1.818400938	**Yes**
		Sickle cell	-0.0211162	0.89					Edu: 0.07		
		Maternal Edu*Sickle	-0.0545878	0.78	No						
	Recent tea	Recent tea	-0.046359	0.72		-0.0273823	0.77	-0.04035	Sickle: 0.67	1.473561388	**Yes**
		Sickle cell	0.2559799	0.25					Tea: 0.31		
		Recent tea*sickle	-0.3603657	0.15	No						
	Recent Sprinkles	Recent Sprinkles	-0.0095748	0.95		-0.0273823	0.77	-0.05344	Sickle: 0.59	1.951607425	**Yes**
		Sickle cell	-0.0730076	0.48					Sprinkles: 0.87		
		Recent Sprinkles*sickle	0.1565181	0.57	No						
	Malaria parasitemia	Malaria parasitemia	1.0325334	**<0.0001**		-0.0273823	0.77	-0.044406	Sickle: 0.58	1.621711836	**Yes**
		Sickle cell	-0.0230801	0.83					Malaria: <0.0001		
		Malaria*sickle	-0.0627688	0.7	No						
	Recent fever	Recent fever	0.6193295	**<0.0001**		-0.0273823	0.77	-0.047971	Sickle: 0.60	1.751898124	Yes
		Sickle cell	-0.0068523	0.95					Fever: <0.0001		
		Recent fever*sickle	-0.0968981	0.59	No						
	Thalassemia	Thal	0.0488362	0.65		-0.0273823	0.77	0.0009974	Sickle: 0.99	-0.036425355	**Yes**
		sickle cell	0.0429467	0.78					Thal: 0.73		
		Thal*Sickle	-0.0843407	0.69	no						
	G6PD	G6PD	0.0537746	0.75		-0.0273823	0.77	-0.015702	Sickle: 0.87	0.573436125	**Yes**
		sickle cell	-0.0035866	0.97					G6PD: 0.90		
		G6PD*sickle	-0.1605769	0.57	no						

Appendix 4d: Assessing confounding and interaction for the model of TfR/SF index and sickle cell

Model	Column1	Independent variables	Beta 1	Independent variable p-	INTERACTION?	Crude B1 (from SL	p-value	Adjusted B	p-value2	10% Rule	CONFOUNDING?
Index and Sickle Cell	Age	Sickle cell	-0.005	0.99		0.09610711	0.36	0.1002725	Sickle: 0.34	1.043341122	No
	n=846	Age	-0.01	**0.03**					Age: 0.03		
		Age*Sickle	0.005	0.66	No						
	Sex	Sickle cell (x1)	-0.1269369	0.42		0.09610711	0.36	0.0970723	Sickle: 0.35	1.010042649	No
	n=846	Sex (ref = female)	0.2834974	0.2					Sex: 0.0001		
		Sex*sickle	0.0596645	0.81	No						
	Inflammation stage	Sickle cell	0.0336797	0.84		0.09610711	0.36	0.0772394	Sickle: 0.45	0.803680394	**Yes**
	n=846	Inflammation	-0.8604861	**<0.0001**					Inflammation: <0.0001		
		Sickle*Inflammation	0.0719402	0.75	No						
	RBP	Sickle cell	0.3126263	0.39		0.09610711	0.36	0.0985839	Sickle: 0.35	1.02577083	No
	n=846	RBP	0.7455788	**<0.0001**					RBP:<0.0001		
		RBP*Sickle cell	-0.2504984	0.51	No						
	Stunting	Sickle	0.1245997	0.3		0.09610711	0.36	0.085082	Sickle: 0.42	0.885283409	**Yes**
	n=842	Stunting	0.1950316	0.07					Stunting: 0.08		
		Stunting*Sickle	-0.1178893	0.63	No						
	Wasting	Wasting	-0.6169429	**0.01**		0.09610711	0.36	0.0914397	Sickle: 0.39	0.951435331	No
	n=844	Sickle cell	0.0660501	0.52					Wasting: 0.04		
		Wasting*Sickle cell	0.9396297	0.15	No						
	Underweight	Underweight	-0.3641696	**0.03**		0.09610711	0.36	0.1047065	Sickle: 0.32	1.089477147	No
	n=843	Sickle cell	-0.0235693	0.81					Under: 0.30		
		Underweight*Sickle cell	0.9260724	**0.01**	**Yes**						
	SES	SES	0.0223394	0.84		0.09610711	0.36	0.1129438	Sickle: 0.29	1.175186727	**Yes**
	n=829	Sickle cell	0.3541854	0.02					SES: 0.29		
		SES*Sickle	-0.6268327	**0.01**	**Yes**						
	Maternal Education	Maternal Education	0.08632551	0.41		0.09610711	0.36	0.1273426	Sickle: 0.23	1.32500686	**Yes**
	n=822	Sickle cell	0.11718329	0.51					Edu: 0.32		
		Maternal Edu*Sickle	0.01933942	0.93	No						
	Recent tea	Recent tea	0.0537327	0.7		0.09610711	0.36	0.1135848	Sickle: 0.28	1.181856681	**Yes**
	n=818	Sickle cell	-0.309569	0.18					Tea: 0.21		
		Recent tea*sickle	0.5145967	0.07	No						
	Recent Sprinkles	Recent Sprinkles	0.0668713	0.62		0.09610711	0.36	0.125786	Sickle: 0.25	1.308810555	**Yes**
	n=823	Sickle cell	0.177014	0.14					Sprinkles: 0.88		
		Recent Sprinkles*sickle	-0.4097537	0.14	No						
	Malaria parasitemia	Malaria parasitemia	-0.6263147	**<0.0001**		0.09610711	0.36	0.1064472	Sickle: 0.29	1.10758923	**Yes**
	n=840	Sickle cell	0.176435	0.18					Malaria: <0.0001		
		Malaria*sickle	-0.205994	0.35	No						
	Recent fever	Recent fever	-0.4679974	**<0.0001**		0.09610711	0.36	0.1177022	Sickle: 0.25	1.224698152	**Yes**
	n=815	Sickle cell	0.097699	0.45					Fever: <0.0001		
		Recent fever*sickle	0.0471384	0.83	No						
	Thalassemia	Thal	0.02870271	0.8		0.09610711	0.36	0.0626111	Sickle: 0.57	0.651472508	**Yes**
	n=814	sickle cell	0.02107215	0.9					Thal: 0.66		
		Thal*Sickle	0.08351589	0.71	no						
	G6PD	G6PD	-0.0835401	0.66		0.09610711	0.36	0.07225	Sickle: 0.49	0.7517654	**Yes**
	n=818	sickle cell	0.066142	0.55					G6PD: 0.71		
		G6PD*sickle	0.0809555	0.79	no						

Appendix 4e: Assessing confounding and interaction for the model of TfR and alpha-thalassemia

Model	Column1	Independent variables	Beta 1	Independent variable p-value	INTERACTION?	Crude B1 (from SLR)	p-value	Adjusted B1	p-value2	10% Rule	CONFOUNDING?
TfR and Thal	Age	Thal	0.0785724	0.07		0.07904985	0.07	0.0795871	thal 0.07	1.006796344	No
	n=814	Age	-0.002333	0.26					Age: 0.37		
		Age*Thal	0.0024867	0.27	No						
	Sex	Thal	0.1170822	0.05		0.07904985	0.07	0.0757804	thal: 0.09	0.958640402	No
	n=815	Sex (ref = female)	0.1781475	**0.001**					Sex: 0.0003		
		Sex*thal	-0.0821255	0.29	No						
	Inflammation stage	Thal	0.05736561	0.25		0.07904985	0.07	0.0832763	thal: 0.04	1.053465377	No
	n=815	Inflammation	0.27172277	**<0.0001**					Inflammation: <0.0001		
		Thal*Inflammation	0.04152767	0.54	No						
	RBP	thal	0.14508	0.25		0.07904985	0.07	0.0778834	thal: 0.07	0.985244121	No
	n=815	RBP	-0.1859886	0.07					RBP:0.004		
		RBP*thal	-0.0784464	0.56	No						
	Stunting	thal	0.07601753	0.11		0.07904985	0.07	0.0791068	thal: 0.07	1.000720811	No
	n=811	Stunting	0.15970403	**0.01**					Stunting: 0.0002		
		Stunting*thal	0.0104343	0.91	No						
	Wasting	Wasting	0.07256661	0.44		0.07904985	0.07	0.0777521	thal: 0.08	0.983582891	No
	n=812	thal	0.07319967	0.11					Wasting: 0.10		
		Wasting*thal	0.12513783	0.41	No						
	Underweight	Underweight	0.15565644	**0.02**		0.07904985	0.07	0.0813143	thal: 0.06	1.028645975	No
	n=813	thal	0.07841383	0.11					Under: 0.005		
		Underweight*thal	0.02415789	0.83	No						
	SES	SES	0.0344258	0.53		0.07904985	0.07	0.0732998	thal: 0.10	0.927260457	No
	n=798	thal	0.1253218	0.03					SES: 0.51		
		SES*thal	-0.1295143	0.1	No						
	Maternal Education	Maternal Education	-0.0597226	0.14		0.07904985	0.07	0.0731665	thal: 0.10	0.925574179	No
	n=790	thal	0.072441	0.1					Edu: 0.14		
		Maternal Edu*thal	0.0356841	0.6	No						
	Recent tea	Recent tea	0.03407835	0.5		0.07904985	0.07	0.0629107	thal: 0.15	0.79583516	**Yes**
	n=787	thal	0.04384887	0.59					Tea: 0.28		
		Recent tea*thal	0.02293223	0.82	No						
	Recent Sprinkles	Recent Sprinkles	-0.0112569	0.88		0.07904985	0.07	0.0721861	thal: 0.10	0.913171878	No
	n=792	thal	0.0707665	0.13					Sprinkles: 0.93		
		Recent Sprinkles*thal	0.0130238	0.9	No						
	Malaria parasitemia	Malaria parasitemia	0.3997291	**<0.0001**		0.07904985	0.07	0.0758503	thal: 0.06	0.959525034	No
	n=809	thal	0.1078169	0.04					Malaria: <0.0001		
		Malaria*thal	-0.0978908	0.21	No						
	Recent fever	Recent fever	0.1510245	**0.01**		0.07904985	0.07	0.074356	thal: 0.09	0.940621646	No
	n=784	thal	0.0870329	0.14					Fever: 0.002		
		Recent fever*thla	-0.0303493	0.73	No						
	Sickle	Thal	0.0775389	0.09		0.07904985	0.07	0.0773835	Sickle: 0.26	0.97892039	No
	n=814	sickle cell	0.0640188	0.39					Thal: 0.08		
		Thal*Sickle	-0.0008249	0.99	no						
	G6PD	G6PD	-0.0387576	0.71		0.07904985	0.07	0.0837648	thal: 0.06	1.059645274	No
	n=792	thal	0.0857817	0.07					G6PD: 0.36		
		G6PD*thal	-0.0305361	0.82	no						

Appendix 4f: Assessing confounding for the model of ZP and G6PD among boys

Model	Column1	Independent variables	Beta 1	Independent variable p-value	INTERACTION?	Crude B1 (from SLR)	p-value	Adjusted B1	p-value2	10% Rule	CONFOUNDING?
ZP and G6PD in boys	Age	G6PD	0.0532741	0.76		-0.2122523	0.01	-0.194472	G6PD: 0.01	0.916230354	No
		Age	-0.0078493	**0.03**					Age: 0.02		
		Age*G6PD	0.00799788	0.19	No						
	Inflammation stage	G6PD	-0.1560089	0.11		-0.2122523	0.01	-0.196011	G6PD: 0.01	0.923482572	No
		Inflammation	0.2590974	**0.0004**					Inflammation: 0.0002		
		G6PD*Inflammation	-0.0747902	0.64	No						
	RBP	G6PD	-0.6100902	0.06		-0.2122523	0.01	-0.174755	G6PD: 0.03	0.823337603	**YES**
		RBP	-0.5065268	**<0.0001**					RBP: <0.0001		
		RBP*G6PD	0.4788511	0.14	No						
	Stunting	G6PD	-0.2119915	0.04		-0.2122523	0.01	-0.218852	G6PD: 0.01	1.031093656	No
		Stunting	0.1203761	0.07					Stunting: 0.06		
		Stunting*G6PD	-0.0180996	0.92	No						
	Wasting	Wasting	0.010579	0.93		-0.2122523	0.01	-0.213034	G6PD: 0.01	1.003680525	No
		G6PD	-0.2130335	0.01					Wasting: 0.93		
		Wasting*G6PD	0	0	No						
	Underweight	Underweight	0.1171441	0.21		-0.2122523	0.01	-0.2372	G6PD: 0.003	1.117538891	**YES**
		G6PD	-0.2017405	0.02					Under: 0.63		
		Underweight*G6PD	-0.0777884	0.79	No						
	SES	SES	-0.0450687	0.44		-0.2122523	0.01	0.0732998	G6PD: 0.10	-0.345342783	No
		G6PD	-0.3089216	0.001					SES: 0.51		
		SES*G6PD	0.2479222	0.1	No						
	Maternal Education	Maternal Education	-0.0168237	0.76		-0.2122523	0.01	-0.232268	G6PD: 0.003	1.094299096	No
		G6PD	-0.3773121	0.01					Edu: 0.92		
		Maternal Edu*G6PD	0.1971827	0.27	No						
	Recent tea	Recent tea	0.0345119	0.62		-0.2122523	0.01	-0.197493	G6PD: 0.02	0.930463415	No
		G6PD	-0.1774126	0.22					Tea: 0.62		
		Recent tea*G6PD	-0.0240581	0.89	No						
	Recent Sprinkles	Recent Sprinkles	0.0513325	0.59		-0.2122523	0.01	-0.234783	G6PD: 0.003	1.106152442	**YES**
		G6PD	-0.2510659	0.005					Sprinkles: 0.49		
		Recent Sprinkles*G6PD	0.1316499	0.26	No						
	Malaria parasitemia	Malaria parasitemia	0.2992484	**<0.0001**		-0.2122523	0.01	-0.188182	G6PD: 0.02	0.886596282	**YES**
		G6PD	-0.1784146	0.06					Malaria: <0.0001		
		Malaria*G6PD	-0.0404106	0.83	No						
	Recent fever	Recent fever	0.1648238	**0.003**		-0.2122523	0.01	-0.206244	G6PD: 0.01	0.971693593	No
		G6PD	-0.1313917	0.17					Fever: 0.005		
		Recent fever*G6PD	-0.2854142	**0.03**	**YES**						
	Sickle	G6PD	-0.1981047	0.03		-0.2122523	0.01	-0.206691	Sickle: 0.35	0.97379675	No
		sickle cell	-0.0637743	0.4					G6PD: 0.01		
		G6PD*Sickle	-0.0335343	0.85	no						
	Thal	G6PD	-0.1884853	0.23		-0.2122523	0.01	-0.213938	thal: 0.67	1.007941963	No
		thal	-0.0239357	0.72					G6PD: 0.01		
		G6PD*thal	-0.0407747	0.82	no						

Appendix 5: Multiple Linear Regression Models

Appendix 5a – Model of ZP and sickle cell disease/trait

The first model had zinc protoporphyrin (ZP) as the dependent variable and sickle cell as the primary variable (sickle cell was dichotomized such that 1 = sickle cell disease or sickle cell trait and 0 = HbAA). We evaluated and confirmed linearity of age and RBP with ZP. Each additional variable was assessed for interaction and confounding. Interaction was tested by adding an interaction term into the linear regression model with only the primary variable and additional variable of interest. Statistical significance of the interaction term was determined using a significance level of 0.05. If no interaction was identified, the interaction term was removed from the model. Confounding was then assessed. Confounding was determined to be present if the crude parameter estimate for sickle cell (from the simple linear regression model with ZP as the outcome) differed by greater than 10% from the adjusted parameter estimate for sickle cell (from the multiple linear regression model with ZP as outcome, sickle cell as the primary variable and the potential confounder as the additional variable). The independent variables that were evaluated for interaction and confounding were age, sex, inflammation, RBP, stunting, wasting, underweight, thalassemia, G6PD deficiency, recent tea consumption, recent Sprinkle use, malaria parasitemia, recent fever, SES, and maternal education. Interaction was found between sickle cell trait/disease and recent Sprinkle use as well as malaria parasitemia. RBP was a statistically significant predictor (p=0.04) in the interaction model so was included in the exhaustive model. By the same criteria, inflammation, stunting and recent fever were also included in the exhaustive model. The independent variables inflammation, stunting, maternal education, tea consumption in previous 24 hours, thalassemia trait or disease, and G6PD deficiency were found to confound the relationship between ZP and sickle cell trait/disease.

Therefore, an exhaustive multivariate linear regression model with the following independent variables was evaluated: sickle cell trait/disease (primary variable), RBP, recent Sprinkle use, interaction term for recent Sprinkle use, malaria parasitemia, interaction term for malaria parasitemia, and as confounders, inflammation, stunting, thalassemia, G6PD deficiency, maternal education and recent tea consumption. We checked for collinearity (defining significant collinearity as a correlation coefficient >0.80) and none was present. The results of this model are shown in Appendix 5a-1. According to this model, sickle cell was not a statistically significant predictor of zinc protoporphyrin, controlling for all the listed factors (p=0.18).

We then attempted to identify a more parsimonious model by removing the least significant independent variable and assessing whether the regression coefficient for sickle cell disease/trait was changed by more than 10%. If it was, then the variable was retained in the model. If the regression coefficient did not change by more than 10%, the variable was dropped from the model. Additionally, interaction terms with a p-value greater than 0.01 were dropped from the model. As a result, the following variables were dropped from the model: maternal education, thalassemia trait/disease, recent Sprinkles, G6PD deficiency, recent fever, RBP, recent tea, stunting and the interaction term for sickle cell

and recent Sprinkles use. The resulting reduced model is present in Appendix 5a-ii. Sickle cell was not a statistically significant predictor of ZP according to this model (p=0.30). We also reran the reduced model adding back in the blood disorders (G6PD deficiency and 3-level thalassemia) and again the findings were unchanged (p=0.42) (Appendix 5a-iii).

Appendix 5a-i: Multiple linear regression analysis for model of zinc protoporphyrin and sickle cell (exhaustive)

Dependent Variable	R^2	Independent Variables	Regression coefficient	p-value
Zinc protoporphyrin	0.14	Intercept	5.1870994	<.0001
n=731		Sickle cell	0.0916049	0.18
		Thalassemia	0.0307251	0.45
		G6PD Deficiency	-0.1005164	0.23
		Sprinkles	0.0533186	0.44
		Malaria parasitemia	0.2734864	**<0.0001**
		Inflammation	0.1577403	**0.002**
		Recent fever	0.0726059	0.12
		Stunted	0.0839593	0.09
		RBP	-0.1295033	0.09
		Maternal Education	0.0222233	0.57
		Recent tea	0.0628249	0.10
		Sickle cell * Sprinkles	-0.2424742	0.03
		Sickle cell * Malaria	-0.2968483	**0.01**

Appendix 5a-ii: Multiple linear regression analysis for model of zinc protoporphyrin and sickle cell (reduced)

Dependent Variable	R^2	Independent Variables	Regression coefficient	p-value
Zinc protoporphyrin	0.12	Intercept	5.1811494	<.0001
n=846		Sickle cell	0.0593273	0.30
		Malaria parasitemia	0.3103068	<.0001
		Inflammation	0.1889414	<0.0001
		Sickle cell * Malaria	-0.2948064	0.004

Appendix 5a-iii: Final model: Add back hemoglobinopathies (thalassemia as 3 level variable)

Dependent Variable	R^2	Independent Variables	Regression coefficient	p-value
Zinc protoporphyrin	0.14	Intercept	5.1737343	<0.0001
n=788		Sickle cell	0.0515609	0.42
		Thalassemia		
		Trait vs Normal	0.0385239	0.35
		Disease vs Normal	0.0206915	0.75
		G6PD Deficiency	-0.0989661	0.17
		Malaria parasitemia	0.3152676	<0.0001
		Inflammation	0.173846	0.00
		Sickle cell * Malaria	-0.2876793	0.007

Appendix 5b – Model of TfR and sickle cell disease/trait

In this model, soluble transferrin receptor (TfR) was the dependent variable and sickle cell disease or trait the primary variable of interest (dichotomized as previously described). Linearity of the continuous variables age and RBP with TfR was confirmed. Interaction and confounding were assessed as previously described. Interaction with sickle cell disease/trait was found with recent sprinkles use and malaria parasitemia. In the interaction models, inflammation (defined as inflammation stages of incubation, early and late convalescence versus reference), RBP, stunting, underweight and recent fever were statistically significant predictors (while their interaction terms were not) and thus were included in the model. Confounders of the relationship between TfR and sickle cell trait/disease included inflammation, stunting, maternal education level, malaria parasitemia, and G6PD deficiency.

Therefore, an exhaustive multivariate linear regression model with the following independent variables was evaluated: sickle cell trait/disease (primary variable), RBP, stunting, underweight, recent fever, recent Sprinkles use, interaction term for recent Sprinkles use, inflammation stage, malaria parasitemia, interaction term for malaria parasitemia, and the remaining confounders that were not yet included in the model, maternal education level and G6PD deficiency. The results of this model are shown in Appendix 5b-i. According to this model, sickle cell was a statistically significant predictor of TfR, controlling for all the listed factors (p=0.03).

We then attempted to identify a more parsimonious model by the method described above. As a result, the following variables were dropped from the model: maternal education, underweight, G6PD deficiency, recent fever, RBP, and recent Sprinkles use, the interaction term for sickle cell and sprinkles, and the interaction term for sickle cell and malaria parasitemia. The resulting reduced model is presented in Appendix 5b-ii**.** In the reduced model, the independent variables of malaria parasitemia, inflammation and stunting remained statistically significant, but sickle cell did not remain a statistically significant predictor of TfR. We also reinserted the blood disorders (G6PD deficiency and three-level thalassemia) into the model, and the conclusion from the reduced model remained unchanged (p=0.38) (Appendix 5b-iii).

Appendix 5b-i: Exhaustive model for TfR and sickle cell

Dependent Variable	R^2	Independent Variables	Regression coefficient	p-value
TfR	0.15	Intercept	2.1618264	<.0001
(n=765)		**Sickle cell**	0.1749922	**0.03**
		G6PD Deficiency	-0.0618716	0.38
		Malaria parasitemia	0.3030367	**<0.0001**
		Inflammation	0.2106102	**<0.0001**
		RBP	0.0943976	0.29
		Recent fever	0.0528883	0.24
		Sprinkles	0.1180371	0.05
		Underweight	0.0128457	0.85
		Stunted	0.119562	**0.01**
		Maternal Education	0.0000683	1.00
		Sickle cell * Sprinkles	-0.270588	0.01
		Sickle cell * Malaria	-0.2549347	0.04

Appendix 5b-ii: Reduced model for TfR and sickle cell

Dependent Variable	R^2	Independent Variables	Regression coefficient	p-value
TfR	0.14	Intercept	2.28617806	<0.0001
(n=836)		Sickle cell	0.05970073	0.26
		Malaria parasitemia	0.26445222	**<0.0001**
		Inflammation	0.20039159	**<0.0001**
		Stunted	0.12696699	**0.003**

Appendix 5b-iii: Final model for TfR and sickle cell

Dependent Variable	R^2	Independent Variables	Regression coefficient	p-value
TfR	0.14	Intercept	2.2558196	<0.0001
(n=781)		Sickle cell	0.0470668	0.38
		Malaria parasitemia	0.2613155	**<0.0001**
		Inflammation	0.194598	**<0.0001**
		Stunted	0.1204108	**0.01**
		G6PD Deficiency	-0.0524198	0.41
		Thalassemia		
		Trait vs Normal	0.0822847	0.05
		Disease vs Normal	0.0628402	0.36

Appendix 5c – Model of SF and sickle cell disease/trait

In this model, serum ferritin (SF) was the dependent variable and sickle cell disease or trait the primary variable of interest (dichotomized as previously described). Linearity of the continuous variables age and RBP with SF was confirmed. Interaction and confounding were assessed as previously described. Interaction was found between sickle cell disease/trait and wasting, underweight, and SES. In the interaction models, age, inflammation, RBP, malaria parasitemia and recent fever were statistically significant predictors (while their interaction terms were not) and thus were included in the model. Confounders of the relationship between SF and sickle cell trait/disease included age, inflammation, RBP, wasting, underweight, SES, maternal education, recent tea consumption, recent Sprinkles use, malaria parasitemia, recent fever, thalassemia and G6PD deficiency.

Therefore, an exhaustive multivariate linear regression model with the following independent variables was evaluated: sickle cell trait/disease (primary variable), wasting, interaction term for wasting, underweight, interaction term for underweight, SES, interaction term for SES, age, inflammation, RBP, malaria parasitemia, recent fever, and the remaining confounders that were not yet included in the model, age, maternal education, recent tea consumption, recent Sprinkles use, thalassemia and G6PD deficiency. The results of this model are shown in Appendix 5c-i. According to this model, sickle cell was not a statistically significant predictor of SF, controlling for all the listed factors ($p=0.39$).

We then attempted to identify a more parsimonious model by the method described previously, however no variables fell out of the model. We did removed the sickle cell interaction term for wasted and underweight according to the $p<0.01$ cut-off for interaction terms. The resulting reduced model is presented in Appendix 5c-ii, and we again concluded that sickle cell disease/trait is not a statistically significant independent predictor of SF, controlling for other factors ($p=0.06$). Finally, we reran the reduced model with thalassemia as a 3-level variables and the conclusions were unchanged ($p=0.05$) (Appendix 5c-iii).

Appendix 5c-i: Exhaustive model for SF and sickle cell

Dependent Variable	R^2	Independent Variables	Regression coefficient	p-value
SF	0.36	Intercept	2.4878589	<0.0001
(n=732)		Sickle cell	-0.087979	0.39
		G6PD Deficiency	0.0537658	0.64
		Thalassemia	0.0426352	0.59
		Malaria parasitemia	0.6041595	**<0.0001**
		Inflammation stage	0.8677133	**<0.0001**
		RBP	0.0491601	0.71
		Recent fever	0.2647346	**0.001**
		Age	0.0078543	0.07
		Recent Sprinkles	0.0502646	0.63
		Recent tea	-0.0838265	0.39
		SES	-0.1293503	0.11
		Underweight	0.2262909	0.08
		Wasted	0.4086209	0.09
		Maternal Education	-0.0062757	0.93
		Sickle cell * Wasted	-0.7935393	0.06
		Sickle cell * Underweight	-0.690555	0.01
		Sickle cell * SES	0.5703863	**0.005**

Appendix 5c-ii: Reduced model for SF and sickle cell

Dependent Variable	R^2	Independent Variables	Regression coefficient	p-value
SF	0.35	Intercept	2.6377118	<0.0001
(n=747)		Sickle cell	-0.2184879	0.06
		G6PD Deficiency	0.0809041	0.46
		Thalassemia	0.0351921	0.66
		Malaria parasitemia	0.603261	**<0.0001**
		Inflammation	0.8894697	**<0.0001**
		Recent fever	0.2684644	**0.001**
		Recent Sprinkles	0.0638153	0.55
		SES	-0.1216026	0.11
		Underweight	0.1151929	0.31
		Sickle cell * SES	0.5836993	**0.003**

Appendix 5c-iii: Final model for SF and sickle cell

Dependent Variable	R^2	Independent Variables	Regression coefficient	p-value
SF	0.35	Intercept	2.6381434	<0.0001
(n=747)		Sickle cell	-0.2231409	0.0545
		G6PD Deficiency	0.0884517	0.42
		Thalassemia		
		Trait vs Normal	0.0123798	0.88
		Disease vs Normal	0.1259445	0.28
		Malaria parasitemia	0.607897	**<0.0001**
		Inflammation	0.8906353	**<0.0001**
		Recent fever	0.2661654	**0.001**
		Recent Sprinkles	0.0603863	0.57
		SES	-0.1237572	0.1
		Underweight	0.1135607	0.32
		Sickle cell * SES	0.5862747	0.003

Appendix 5d – Model of Index and sickle cell disease/trait

In this model, TfR/Ferritin index was the dependent variable and sickle cell disease or trait the primary variable of interest (dichotomized as previously described). Linearity of the continuous variables age and RBP with index was confirmed. Interaction and confounding were assessed as previously described. Interaction was found between sickle cell disease/trait and underweight and SES. In the interaction models, age, inflammation, RBP, wasting, malaria parasitemia and recent fever were statistically significant predictors (while their interaction terms were not) and thus were included in the model. Confounders of the relationship between index and sickle cell trait/disease included inflammation, stunting, SES, maternal education, recent tea consumption, recent Sprinkles use, malaria parasitemia, recent fever, alpha-thalassemia (trait or disease), and G6PD deficiency.

Therefore, an exhaustive multivariate linear regression model with the following independent variables was evaluated: sickle cell trait/disease (primary variable), underweight, interaction term for underweight, SES, interaction term for SES, age, inflammation, RBP, wasting, malaria parasitemia, recent fever, and the remaining confounders that were not yet included in the model, maternal education, recent tea consumption, recent Sprinkles use, thalassemia and G6PD deficiency. The results of this model are shown in Appendix 5d-i. According to this model, sickle cell was not a statistically significant predictor of SF, controlling for all the listed factors (p=0.26).

We then attempted to identify a more parsimonious model by the method described previously. As a result, recent Sprinkles, maternal education, RBP, wasting, recent tea consumption and the interaction term for sickle cell and SES were dropped from the model. The resulting reduced model is presented in Appendix 5d-ii, and we again concluded that sickle cell disease/trait is not a statistically significant independent predictor of TfR/SF index, controlling for other factors (p=0.26). When thalassemia is taken as a three-level variables, we again come to the same conclusion that sickle cell is not an independent predictor of TfR/SF index (p=0.30) (Appendix 5d-iii).

Appendix 5d-i: Exhaustive model for TfR/Ferritin Index and sickle cell

Dependent Variable	R^2	Independent Variables	Regression coefficient	p-value
Index	0.19	Intercept	6.6033873	<0.0001
(n=730)		Sickle cell	0.1417908	0.26
		G6PD Deficiency	-0.0883098	0.55
		Thalassemia	0.0386218	0.69
		Malaria parasitemia	-0.3392475	**0.004**
		Inflammation	-0.6480085	**<0.0001**
		RBP	0.0681487	0.68
		Recent fever	-0.232497	**0.03**
		Age	-0.0158926	**0.001**
		Recent Sprinkles	0.0201492	0.87
		Recent tea	0.1689113	0.14
		SES	0.0527259	0.61
		Stunted	0.3878338	**0.001**
		Underweight	-0.4724591	**0.01**
		Wasted	-0.1742515	0.5
		Maternal Education	0.0181318	0.83
		Sickle cell * Underweight	0.9563259	**0.0049**
		Sickle cell * SES	-0.62487	0.0129

Appendix 5d-ii: Reduced model for TfR/Ferritin Index and sickle cell

Dependent Variable	R^2	Independent Variables	Regression coefficient	p-value
Index	0.18	Intercept	6.842021	<0.0001
(n=750)		Sickle cell	-0.1005321	0.26
		G6PD Deficiency	-0.1371291	0.34
		Thalassemia	0.0581467	0.54
		Malaria parasitemia	-0.3435133	0.003
		Inflammation	-0.6748454	<0.0001
		Recent fever	-0.2526504	0.01
		Age	-0.014229	0.003
		SES	-0.0861351	0.32
		Stunted	0.4196119	0.001
		Underweight	-0.5246657	0.003
		Sickle cell * Underweight	0.9283479	0.007

Appendix 5d-iii: Final model for TfR/Ferritin Index and sickle cell

Dependent Variable	R^2	Independent Variables	Regression coefficient	p-value
Index	0.18	Intercept	6.8413123	<0.0001
(n=750)		Sickle cell	-0.0955682	0.30
		G6PD Deficiency	-0.1452547	0.32
		Thalassemia		
		Trait vs Normal	0.0826434	0.42
		Disease vs Normal	-0.0401504	0.78
		Malaria parasitemia	-0.348563	0.002
		Inflammation	-0.6762458	<0.0001
		Recent fever	-0.2502798	0.01
		Age	-0.0141758	0.003
		SES	-0.0842037	0.33
		Stunted	0.4178996	0.001
		Underweight	-0.5204461	0.003
		Sickle cell * Underweight	0.9221589	0.007

Appendix 5e – Model of TfR and alpha-thalassemia

In this model, TfR was the dependent variable and alpha thalassemia disease or trait the primary variable of interest. Linearity of the continuous variables age and RBP with TfR was confirmed. Interaction and confounding were assessed as previously described. No interaction terms were statistically significant. In the interaction models, sex, inflammation, stunting, underweight, malaria parasitemia and recent fever were statistically significant predictors (while their interaction terms were not) and thus were included in the model. Recent tea consumption was the only confounder of the relationship between TfR and thalassemia included maternal education.

Therefore, an exhaustive multivariate linear regression model with the following independent variables was evaluated: alpha thalassemia trait/disease (primary variable), sex, inflammation, stunting, underweight, malaria parasitemia, recent fever, and the confounder recent tea consumption. The results of this model are shown in Appendix 5e-i. According to this model, sickle cell was not a statistically significant predictor of TfR, controlling for all the listed factors (p=0.09).

We then attempted to identify a more parsimonious model by the method described previously. As a result, underweight was dropped from the model. The resulting reduced model is presented in Appendix 5e-ii, and we concluded that thalassemia is not a statistically significant independent predictor of TfR, controlling for other factors (p=0.09). There was no statistically significant collinearity (all VIF < 10).

We also reran the model reinserting the other blood disorders (sickle cell and G6PD deficiency) and using a 3-level variable for thalassemia (normal [reference], trait, and disease). Again, we found that thalassemia is not a statistically significant predictor of TfR, adjusting for other factors, however the finding was close to statistical significance for thalassemia trait (p=0.06) (Appendix 5e-iii).

Appendix 5e-i: Exhaustive model of TfR and alpha-thalassemia

Dependent Variable	R^2	Independent Variables	Regression coefficient	p-value
TfR	0.16	Intercept	2.2912348	<0.0001
(n=767)		Thalassemia	0.0679968	0.09
		Sex	-0.1460586	**<0.0001**
		Malaria parasitemia	0.248287	**<0.0001**
		Inflammation	0.1985921	**<0.0001**
		Recent fever	0.0388183	0.36
		Recent tea	0.0490212	0.2
		Stunted	0.0934804	0.05
		Underweight	0.0244268	0.72

Appendix 5e-ii: Reduced model of TfR and alpha-thalassemia

Dependent Variable	R^2	Independent Variables	Regression coefficient	p-value
TfR	0.16	Intercept	2.2905482	<0.0001
(n=767)		Thalassemia	0.0678107	0.09
		Sex	-0.1453766	**<0.0001**
		Malaria parasitemia	0.2482776	**<0.0001**
		Inflammation	0.1998083	**<0.0001**
		Recent fever	0.0392868	0.36
		Recent tea	0.0492345	0.2
		Stunted	0.1012113	**0.02**

Appendix 5e-iii: Final model of TfR and alpha-thalassemia

Dependent Variable	R^2	Independent Variables	Regression coefficient	p-value
TfR	0.16	Intercept	2.2698646	<0.0001
(n=744)		Thalassemia		
		Trait vs Normal	0.0799065	0.06
		Disease vs Normal	0.0237969	0.72
		Sickle Cell	0.0534983	0.3
		G6PD Deficiency	-0.0573412	0.43
		Sex	-0.1467714	**<0.0001**
		Malaria parasitemia	0.2419251	**<0.0001**
		Inflammation	0.2048285	**<0.0001**
		Recent fever	0.0437909	0.33
		Recent tea	0.0668627	0.1
		Stunted	0.092604	**0.04**

Appendix 5f – Model of ZP and G6PD among boys

In this model, ZP was the dependent variable and G6PD deficiency was the primary variable of interest. Linearity of the continuous variables age and RBP with ZP was confirmed. Interaction and confounding were assessed as previously described. The interaction term for G6PD deficiency and recent fever was statistically significant. In the interaction models, age, inflammation status, RBP, and malaria parasitemia were statistically significant predictors (while their interaction terms were not) and thus were included in the model. The confounders of the relationship between ZP and G6PD deficiency among boys included RBP, underweight, recent Sprinkles use, and malaria parasitemia.

Therefore, an exhaustive multivariate linear regression model with the following independent variables was evaluated: G6PD deficiency (primary variable), age, RBP, underweight, recent fever, inflammation status, malaria parasitemia, recent Sprinkles use, and an interaction term for G6PD deficiency and recent fever. The results of this model are shown in Appendix 5f-i. According to this model, G6PD deficiency was not a statistically significant predictor of ZP, controlling for all the listed factors (p=0.18).

We then attempted to identify a more parsimonious model by the method described previously. As a result, only the interaction term was dropped from the model. The resulting reduced model is presented in Appendix 5f-ii, and we concluded that G6PD deficiency is a statistically significant independent predictor of ZP, controlling for other factors (p=0.03). There was no statistically significant collinearity (all VIF < 10).

We also reran the model reinserting the other blood disorders (sickle cell and alpha-thalassemia as a 3-level variable). Again, we found that G6PD deficiency in boys is a statistically significant predictor of ZP, adjusting for other factors (p=0.04) (Appendix 5f-iii).

Appendix 5f-i: Exhaustive model of ZP and G6PD deficiency

Dependent Variable	R^2	Independent Variables	Regression coefficient	p-value
ZP	0.13	Intercept	5.70	<0.0001
(n=390)		G6PD Def	-0.13	0.18
		Age	-0.01	**0.02**
		RBP	-0.26	**0.03**
		Malaria parasitemia	0.21	**0.004**
		Inflammation	0.11	0.14
		Recent fever	0.06	0.31
		Recent Sprinkles	0.07	0.44
		Underweight	0.02	0.78
		G6PD Def*Recent fever	-0.15	0.3

Appendix 5f-ii: Reduced model of ZP and G6PD deficiency

Dependent Variable	R^2	Independent Variables	Regression coefficient	p-value
ZP	0.13	Intercept	5.70	<0.0001
(n=390)		**G6PD Def**	**-0.17**	**0.03**
		Age	**-0.01**	**0.02**
		RBP	**-0.26**	**0.03**
		Malaria parasitemia	**0.21**	**0.002**
		Inflammation	0.11	0.14
		Recent fever	0.05	0.37
		Recent Sprinkles	0.08	0.41
		Underweight	0.03	0.75

Appendix 5f-iii: Final model of ZP and G6PD deficiency

Dependent Variable	R^2	Independent Variables	Regression coefficient	p-value
ZP	0.13	Intercept	5.7047734	<0.0001
(n=375)		**G6PD Def**	-0.168613	**0.04**
		Sickle Cell	-0.0811008	**0.24**
		Thalassemia		
		Trait vs Normal	0.0160794	**0.8**
		Disease vs Normal	-0.055114	**0.64**
		Age	-0.0078072	**0.03**
		RBP	-0.245273	**0.03**
		Malaria parasitemia	0.2084478	**0.01**
		Inflammation	0.0933665	0.21
		Recent fever	0.0809973	0.19
		Recent Sprinkles	0.0761536	0.41
		Underweight	0.0073915	0.93

Appendix 6: Interaction terms

For the statistically significant interaction terms, we evaluated the mean iron indicator by the 4 levels of the interaction term to further understand this relationship.

For the ZP and sickle cell model, we evaluated the mean ZP for each subgroup using PROC SURVEYMEANS (Appendix 6a). Investigating the interaction between sickle cell and malaria parasitemia, we found, predictably, that children without sickle cell who did not have malaria parasitemia had the lowest mean ZP. However, the subgroup with the highest mean ZP was the children without sickle cell who had malaria parasitemia (this value was higher than the mean ZP for children *with* sickle cell trait/disease and malaria parasitemia).

Appendix 6a: Mean ZP by sickle cell and malaria parasitemia status

Category	n	Mean ZP	95% CI
Sickle cell + / Malaria parasitemia	54	224.40	200.4 - 251.3
Sickle cell + / No malaria parasitemia	105	206.16	179.9 - 236.2
Sicle cell - / Malaria parasitemia	221	287.11	271.0 - 304.2
Sickle cell - / No malaria parasitemia	466	195.19	183.5 - 207.6

For the statistically significant interaction term in the SF and sickle cell model, we evaluated the mean SF for each subgroup using PROC SURVEYMEANS (Appendix 6b). Investigating the interaction between sickle cell and poverty, we found that children with sickle cell disease or trait who are also poor have, counter-intuitively, the highest mean SF, while children with sickle cell disease who are not poor have the lowest mean SF. The mean SF in children without sickle cell disease or trait was minimally different between those who were poor and those who were not poor.

Appendix 6b: Mean SF by sickle cell and poverty status

Category	n	Mean SF	95% CI
Sickle cell + / Poor	59	44.7	34.1-58.7
Sickle cell + / Not poor	96	26.1	20.4-33.4
Sicle cell - / Poor	272	32.8	27.8-38.8
Sickle cell - / Not poor	402	33.9	29.3-39.1

For the statistically significant interaction term in the Index and sickle cell model, we evaluated the mean TfR/SF index for each subgroup using PROC SURVEYMEANS (Appendix 6c). Investigating the interaction between sickle cell and underweight, we found, predictably, that children with sickle cell who were underweight had the highest index value. However, the subgroup with the lowest index value was the children without sickle cell disease or trait who were underweight.

Appendix 6c : Mean TfR/SF index by sickle cell and underweight status

Category	n	Mean TfR/SF index	95% CI
Sickle cell + / Underweight	23	672.8	357.7 - 1265.6
Sickle cell + / Not underweight	135	383.6	320.5 - 459.2
Sicle cell - / Underweight	78	272.9	197.7 - 376.7
Sickle cell - / Not underweight	608	392.7	351.1 - 439.3

Appendix 7: SAS Code

1. 'thesis_datasteps_aug19'

```
*********************************;
* Kiersten Derby
* Thesis Datasets
* Feb 6, 2012
*
* This file contains the datasteps
to create the data file from which
the rest of the analysis is based
*********************************;

libname thesis 'H:\Thesis';
proc import datafile = 'H:\Thesis\Merged FU2010 database REV_19AUG12'
out=thesis dbms = sav replace;
run;
proc contents data=thesis;
run;
proc freq data=thesis;
      tables MomEduc_2;
run;

data samplesize (drop=cdob EntryDate DateSlideTaken DateSlideRead DATE
MomName); *remove identifiers;
      set work.thesis;
      if Hb ne '.'; *Only include if Hb measured;
      if age_6to35 ne '0'; *Only include if age 6-35 mo;
      if (meanzp ne '.' or sf ne '.' or tfr ne '.'); *Only include if they
have at least one measure of iron status;

      if crp >10 then crp10=1;
      if crp ne '.' and crp <=10 then crp10=0;

      *log transform non-gaussian variables;
      log_sf = log(sf);
      log_zp = log(meanzp);
      log_tfr = log(tfr);
      log_index = log(SF_index);
      log_crp = log(crp);

      *create dummy variables for sex;
      if sex=1 then sex1 = 0; *male;
      if sex=2 then sex1 = 1; *female;

      *create group-specific correction factors using Grant et als protocol;
      if (crp le 5 and agp le 1) then cf=1; *reference;
      if (crp > 5 and agp le 1) then cf=2; *incubation;
      if (crp > 5 and agp >1) then cf=3; *early convalescence;
      if (crp le 5 and agp >1) then cf=4; *late convalescence;
      if (crp =. or agp =.) then cf=.;
```

```
*log convert for geometric means;
if cf=1 then log_ref_sf=log(sf);
if cf=1 then log_ref_tfr=log(tfr);
if cf=1 then log_ref_zp=log(meanzp);
if cf=1 then log_ref_index=log(sf_index);
if cf=1 then log_ref_hb=log(hb);

if cf=2 then log_inc_sf=log(sf);
if cf=2 then log_inc_tfr=log(tfr);
if cf=2 then log_inc_zp=log(meanzp);
if cf=2 then log_inc_index=log(sf_index);
if cf=2 then log_inc_hb=log(hb);

if cf=3 then log_early_sf=log(sf);
if cf=3 then log_early_tfr=log(tfr);
if cf=3 then log_early_zp=log(meanzp);
if cf=3 then log_early_index=log(sf_index);
if cf=3 then log_early_hb=log(hb);

if cf=4 then log_late_sf=log(sf);
if cf=4 then log_late_tfr=log(tfr);
if cf=4 then log_late_zp=log(meanzp);
if cf=4 then log_late_index=log(sf_index);
if cf=4 then log_late_hb=log(hb);

*apply correction factors to iron indicators;
if cf=1 then cf_sf=sf;
if cf=1 then cf_tfr=tfr;
if cf=1 then cf_zp=meanzp;
if cf=1 then cf_index=sf_index;
if cf=1 then cf_hb=hb;

if cf=2 then cf_sf=sf*0.71;
if cf=2 then cf_tfr=tfr*1.12;
if cf=2 then cf_zp=meanzp*0.97;
if cf=2 then cf_index=sf_index*1.58;
if cf=2 then cf_hb=hb*0.97;

if cf=3 then cf_sf=sf*0.21;
if cf=3 then cf_tfr=tfr*0.71;
if cf=3 then cf_zp=meanzp*0.68;
if cf=3 then cf_index=sf_index*3.34;
if cf=3 then cf_hb=hb*1.25;

if cf=4 then cf_sf=sf*0.50;
if cf=4 then cf_tfr=tfr*0.77;
if cf=4 then cf_zp=meanzp*0.84;
if cf=4 then cf_index=sf_index*1.54;
if cf=4 then cf_hb=hb*1.12;

*apply cut-offs;
if 0<cf_sf<12 and crp<10 then cf_lowsf=1;
if cf_sf>=12 and crp<10 then cf_lowsf=0;
```

```
      if cf_tfr>8.3 then cf_elevatedtfr=1;
      if 0<cf_tfr<=8.3 then cf_elevatedtfr=0;

      if cf_zp>80 then cf_elevatedzp=1;
      if 0<cf_zp<=80 then cf_elevatedzp=0;

      if cf_index>500 then cf_elevatedindex=1;
      if 0<cf_index<=500 then cf_elevatedindex=0;

      if 0<cf_hb<11.0 then cf_anemic=1;
      if cf_hb>=11.0 then cf_anemic=0;

      *create dummy variables for Correction Factor groupings;
      if cf = 1 then reference=1;
            else reference=0;
      if cf = 2 then incubation=1;
            else incubation=0;
      if cf = 3 then early_conv=1;
            else early_conv=0;
      if cf = 4 then late_conv=1;
            else late_conv = 0;

      *create variable called inflammation1 that indicates any versus no
inflammation;
      if cf = 2 or cf = 3 or cf = 4 then inflammation1 = 1;
            else inflammation1 = 0;

      *create numeric variables for MomEduc_2 where 0=less than completed
primary school and 1=completed primary school or greater;
      if MomEduc_2 = "1-None" or MomEduc_2 = "2-Some Primary School" then
education = 0;
      if MomEduc_2 = "3-Completed Primary School" or MomEduc_2 = "4-Some
Secondary School" or MomEduc_2 = "5-Completed Secondary School"
            or MomEduc_2 = "6-Any Trade School or Uni" then education = 1;
      if MomEduc_2 = " " or MomEduc_2 = "88-Other" then education = .;

      *Create numeric variables for tea consumption in last 24 hours;
      if TeaYest = "0-No" then tea24=0;
      if TeaYest = "1-Yes" then tea24=1;
      if TeaYest = "99-Don't know" then tea24=.;

      *Create numeric variable for sprinkle use in last 24 hours;
      if SprkUseYest = "0-No" then sprinkles24 = 0;
      if SprkUseYest = "1-Yes" then sprinkles24 = 1;
      if SprkUseYest = " " then sprinkles24 = .;

      *Create numeric variables for fever in last 24 hours;
      if fever24h = "0-No" then fever24 = 0;
      if fever24h = "1-Yes" then fever24 = 1;
      if fever24h = "99-Don't know" or fever24h = " " then fever24 = .;

      int_inc_sickle = incubation*sickleYN;
      int_sprinkles_sickle = sprinkles24*sickleYN;
      int_posmalaria_sickle = posmalaria*sickleYN;
```

```
run;

proc freq data=work.samplesize;
      tables MomEduc_2*edu;
run;


proc format;
      value sickleYN
            0 = "1 No Sickle"
            1 = "0 Yes Sickle"
      ; *No sickle as reference;

      value cf
            1 = "4 Reference"
            2 = "1 Incubation"
            3 = "2 Early Convalescence"
            4 = "3 Late Convalescence"
      ; *Reference as reference;

      value stunted
            0 = "1 Not stunted"
            1 = "0 Stunted"
      ; *No stunting as reference;

      value wasted
            0 = "1 Not wasted"
            1 = "0 Wasted"
      ; *No wasting as reference;

      value underwt
            0 = "1 Not underweight"
            1 = "0 Underweight"
      ; *Not underweight as reference;

      value poor
            0 = "1 Not poor"
            1 = "0 Poor"
      ; *Not poor as reference;

      value yesno
            0 = "2 No"
            1 = "1 Yes"
            . = "0 Missing"
      ; *NO as reference;

      value thal
            1 = "3 Normal"
            2 = "1 Trait"
            3 = "2 Disease"
      ; *Reference as reference;

run;
```

```
/**Check new variables;*/
/*proc freq data=work.samplesize;*/
/*    tables MomEduc_2*education teayest*tea24 sprkuseyest*sprinkles24
fever24h*fever24 / nocum nopercent;*/
/*run;*/

*End up with n=854 given above exclusions;
proc contents data=work.samplesize;
run;
```

2. 'thesis_regression_aug19'

This file contains the code for the descriptive statistics presented in Table 1 as well as the regression analysis for the model of ZP and sickle cell.

```
*******************************;
* Kiersten Derby                *;
* Thesis Regression Analysis  *;
* Feb 6, 2012                                *;
* This file contains descriptive statistics (information for table 1)
  and regression analysis for ZP and sickle cell

*******************************;

*-------------------------------------------------------------;
*Call in 'samplesize' data from thesis_datasteps.sas    ;
*-------------------------------------------------------------;

%include "H:\Thesis\thesis_datasteps_aug19.sas";

*Set reference values;
proc format;
      value thalYN
            0 = "1 No Thal"
            1 = "0 Yes Thal"
      ;
      value sickleYN
            0 = "1 No Sickle"
            1 = "0 Yes Sickle"
      ; *No sickle as reference;

      value cf
            1 = "4 Reference"
            2 = "1 Incubation"
            3 = "2 Early Convalescence"
            4 = "3 Late Convalescence"
      ; *Reference as reference;

      value stunted
            0 = "1 Not stunted"
            1 = "0 Stunted"
```

```
      ; *No stunting as reference;

      value wasted
            0 = "1 Not wasted"
            1 = "0 Wasted"
      ; *No wasting as reference;

      value underwt
            0 = "1 Not underweight"
            1 = "0 Underweight"
      ; *Not underweight as reference;

      value poor
            0 = "1 Not poor"
            1 = "0 Poor"
      ; *Not poor as reference;

      value yesno
            0 = "2 No"
            1 = "1 Yes"
            . = "0 Missing"
      ; *NO as reference;

      value sickle
            1 = "3 Normal"
            2 = "2 Trait"
            3 = "1 Disease"
            . = "0 Missing"
      ; *normal as reference;

run;

*Table 1: Descriptive statistics for categorical variables;
proc freq data=samplesize;
      tables sex anemic lowsf elevatedzp elevatedtfr elevatedSF_index lowiron
crp10 posmalaria
                  sickle_cl thal hapto g6pd_def stunted wasted underwt
elevatedcrp elevatedagp clinical_malaria
                  poor education teayest sprkuseyest lowrbp / binomial
(p=0.05);
      exact binomial;
run;

**use surveyfreq with cluster design for descriptive statistics**;
proc surveyfreq data=samplesize;
      cluster cluster;
      tables sex anemic lowsf elevatedzp elevatedtfr elevatedSF_index lowiron
crp10 posmalaria
                  sickle_cl thal hapto g6pd_def stunted wasted underwt
elevatedcrp elevatedagp clinical_malaria
                  poor education teayest sprkuseyest lowrbp inflammation1
fever24/ cl;
run;
```

```
proc sort data=samplesize;
      by sickle_cl;
run;
proc surveyfreq data=samplesize;
      cluster cluster;
      by sickle_cl;
      tables sickle_cl / cl;
run;

*Table 1: Descriptive statistics for Continuous variables - non log-
transformed;
proc univariate data=samplesize PLOT cibasic;
      var childagemonths_lab hb sf meanzp tfr SF_index ironstores crp agp
rbp;
      histogram childagemonths_lab hb sf meanzp tfr SF_index ironstores crp
agp rbp / NORMAL;
run;

*Table 1: Descriptive statistics for Continuous variables - log transformed
if non-gaussian;
proc univariate data=samplesize PLOT cibasic;
      var childagemonths_lab hb log_sf log_zp log_tfr log_index log_crp agp ;
      histogram childagemonths_lab hb log_sf log_zp log_tfr log_index log_crp
agp / NORMAL;
run;

*create new rbp variable that identifies those that are vit A deficient;
data rbp;
      set work.samplesize;
      if rbp < 0.70 then lowrbp =1;
      else if rbp >= 0.70 then lowrbp = 0;
run;
proc surveyfreq data=work.rbp;
      cluster cluster;
      tables lowrbp / cl;
run;

*create new sf variable that identifies those that are iron deficient;
data sf;
      set work.samplesize;
      if log_sf < 2.48490665 then lowferr = 1;
      else if log_sf >= 2.48490665 then lowferr = 0;
run;
proc surveyfreq data=work.sf;
      cluster cluster;
      tables lowferr / cl;
run;

*------------------------------------------------------------;
*Regression Analysis: check for interaction and confounding   ;
*------------------------------------------------------------;
**********************;
* ZP and Sickle       *;
**********************;
```

```
***Two ways to assess for confounding - compare Beta1 and compare difference
in mean
between Sickle (trait and disease) and No Sickle;

*find crude difference in mean ZP between sickle cell (trait and disease) and
non-sickle cell;
proc sort data=samplesize;
      by sickleYN;
run;
proc univariate data=samplesize;
      var log_zp;
      by sickleYN;
run;

*-----;
* AGE ;
*-----;
*Interaction: Test for interaction and confounding between sickleYN and age;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = sickleYN childagemonths_lab sickleYN*childagemonths_lab
/ solution;
      format sickleYN sickleYN.;
run;

*confounding: method 1 (means);
*No interaction between ZP and age - now look for confounding (without
interaction term);
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = sickleYN childagemonths_lab / solution;
      lsmeans sickleYN / diff;
run;

*confounding: method 2 (betas);
*SLR for crude beta;
proc surveyreg data=samplesize;
      class sickleYN;
      cluster cluster;
      model log_zp = sickleYN / solution;
      format sickleYN sickleYN.;
run;

*MLR for adjusted beta;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = sickleYN childagemonths_lab / solution;
      format sickleYN sickleYN.;
run;


*-----;
* SEX ;
```

```
*-----;
*Interaction: Test for interaction and confounding between sickleYN and sex;
proc surveyreg data=samplesize;
      class sickleYN sex1;
      model log_zp = sickleYN sex1 sickleYN*sex1 / solution;
      cluster cluster;
run;
*no interaction between sickle cell and sex - now look for confounding
without interaction term in model;

*Confounding: method 1 (means);
proc surveyreg data=samplesize;
      class sickleYN sex1;
      cluster cluster;
      model log_zp = sickleYN sex1 / solution;
      lsmeans sickleYN / diff;
run;

*confounding: method 2 (betas);
*SLR for crude beta - same as above: beta = -0.04340;

*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN sex1; *reference sex is 1 (female);
      cluster cluster;
      model log_zp = sickleYN sex1 / solution;
      format sickleYN sickleYN.;
run;

*---------;
* Inflamm ;
*---------;
*Inflammation: Test for interaction and confounding between sickleYN and
inflammation;
proc surveyreg data=samplesize;
      class sickleYN cf;
      model log_zp = sickleYN cf sickleYN*cf / solution;
      cluster cluster;
      format sickleYN sickleYN. cf cf.;
run;

*use 'inflammation1' instead of 'cf';
proc surveyreg data=samplesize;
      model log_zp = sickleYN inflammation1 sickleYN*inflammation1 /
solution;
      cluster cluster;
      format sickleYN sickleYN. inflammation1 yesno.;
run;

*Confounding: method 1;
proc surveyreg data=samplesize;
      class sickleYN cf;
      cluster cluster;
      model log_zp = sickleYN cf / solution;
```

```
      lsmeans sickleYN / diff;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = sickleYN inflammation1 / solution;
      format sickleYN sickleYN. inflammation1 yesno.;
run;

*----;
*RBP ;
*----;
*Interaction: Test for interaction and confounding between sickleYN and RBP;
proc surveyreg data=samplesize;
      class sickleYN;
      model log_zp = sickleYN rbp sickleYN*rbp / solution;
      cluster cluster;
      format sickleYN sickleYN.;
run;

*Confounding: method 1;
proc surveyreg data=samplesize;
      class sickleYN;
      cluster cluster;
      model log_zp = sickleYN rbp / solution;
      lsmeans sickleYN / diff;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN;
      cluster cluster;
      model log_zp = sickleYN rbp / solution;
      format sickleYN sickleYN.;
run;


*---------;
*stunting ;
*---------;
*Interaction: Test for interaction and confounding between sickleYN and
stunting;
proc surveyreg data=samplesize;
      class sickleYN stunted;
      model log_zp = sickleYN stunted sickleYN*stunted / solution;
      cluster cluster;
      format sickleYN sickleYN. stunted stunted.;
run;

*confounding: method 1;
proc surveyreg data=samplesize;
```

```
      class sickleYN stunted;
      cluster cluster;
      model log_zp = sickleYN stunted / solution;
      lsmeans sickleYN / diff;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN stunted;
      cluster cluster;
      model log_zp = sickleYN stunted / solution;
      format sickleYN sickleYN. stunted stunted.;
run;


*---------;
*Wasting  ;
*---------;
*Interaction: Test for interaction and confounding between sickleYN and
stunting;
proc surveyreg data=samplesize;
      class sickleYN wasted;
      model log_zp = sickleYN wasted sickleYN*wasted / solution;
      cluster cluster;
      format sickleYN sickleYN. wasted wasted.;
run;

*Confounding: method 1;
proc surveyreg data=samplesize;
      class sickleYN wasted;
      cluster cluster;
      model log_zp = sickleYN wasted / solution;
      lsmeans sickleYN / diff;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN wasted;
      cluster cluster;
      model log_zp = sickleYN wasted / solution;
      format sickleYN sickleYN. wasted wasted.;
run;

*-------------;
*Underweight  ;
*-------------;
*Interaction: Test for interaction and confounding between sickleYN and
stunting;
proc surveyreg data=samplesize;
      class sickleYN underwt;
      model log_zp = sickleYN underwt sickleYN*underwt / solution;
      cluster cluster;
```

```
      format sickleYN sickleYN. underwt underwt.;
run;

*Confounding: method 1;
proc surveyreg data=samplesize;
      class sickleYN underwt;
      cluster cluster;
      model log_zp = sickleYN underwt / solution;
      lsmeans sickleYN / diff;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN underwt;
      cluster cluster;
      model log_zp = sickleYN underwt / solution;
      format sickleYN sickleYN. underwt underwt.;
run;

*-------------;
* SES         ;
*-------------;
proc freq data=samplesize;
      tables poor;
run;
*SES: Test for interaction and confounding between sickleYN and SES where 1=
quintiles 1 or 2, 0 = quintiles 3-5;
*Interaction;
proc surveyreg data=samplesize;
      class sickleYN poor;
      model log_zp = sickleYN poor sickleYN*poor / solution;
      cluster cluster;
      format sickleYN sickleYN. poor poor.;
run;

*Confounding: method 1;
proc surveyreg data=samplesize;
      class sickleYN poor;
      cluster cluster;
      model log_zp = sickleYN poor / solution;
      lsmeans sickleYN / diff;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN poor;
      cluster cluster;
      model log_zp = sickleYN poor / solution;
      format sickleYN sickleYN. poor poor.;
run;

*-------------;
```

```
* maternal edu;
*-------------;

*Maternal Education: Test for interaction and confounding between sickleYN
and education;
*Interaction;
proc surveyreg data=samplesize;
      model log_zp = sickleYN education sickleYN*education / solution;
      cluster cluster;
      format sickleYN sickleYN. education yesno.;
run;

*confounding: method 1;
proc surveyreg data=samplesize;
      class sickleYN education;
      cluster cluster;
      model log_zp = sickleYN education / solution;
      lsmeans sickleYN / diff;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = sickleYN education / solution;
      format sickleYN sickleYN. education yesno.;
run;

*-------------;
* recent tea  ;
*-------------;

*Interaction: Test for interaction and confounding between sickleYN and tea;
proc surveyreg data=samplesize;
      class sickleYN tea24;
      model log_zp = sickleYN tea24 sickleYN*tea24 / solution;
      cluster cluster;
      format sickleYN sickleYN. tea24 yesno.;
run;

*Confounding:method 1;
proc surveyreg data=samplesize;
      class sickleYN tea24;
      cluster cluster;
      model log_zp = sickleYN tea24 / solution;
      lsmeans sickleYN / diff;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN tea24;
      cluster cluster;
      model log_zp = sickleYN tea24 / solution;
```

```
      format sickleYN sickleYN. tea24 yesno.;
run;

*-------------;
* Sprinkles;
*-------------;

*Interaction: Test for interaction and confounding between sickleYN and
sprinkles;
proc surveyreg data=samplesize;
      class sickleYN sprinkles24;
      model log_zp = sickleYN sprinkles24 sickleYN*sprinkles24 / solution;
      cluster cluster;
      format sickleYN sickleYN. sprinkles24 yesno.;
run;

*confounding: method 1;
proc surveyreg data=samplesize;
      class sickleYN sprinkles24;
      cluster cluster;
      model log_zp = sickleYN sprinkles24 / solution;
      lsmeans sickleYN / diff;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN sprinkles24;
      cluster cluster;
      model log_zp = sickleYN sprinkles24 / solution;
      format sickleYN sickleYN. sprinkles24 yesno.;
run;

*-------------;
* Malaria     ;
*-------------;

*Interaction: Test for interaction and confounding between sickleYN and
malaria;
proc surveyreg data=samplesize;
      class sickleYN posmalaria;
      model log_zp = sickleYN posmalaria sickleYN*posmalaria / solution;
      cluster cluster;
      format sickleYN sickleYN. posmalaria yesno.;
run;

*confounding: method 1;
proc surveyreg data=samplesize;
      class sickleYN posmalaria;
      cluster cluster;
      model log_zp = sickleYN posmalaria / solution;
      lsmeans sickleYN / diff;
run;
```

```
*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN posmalaria;
      cluster cluster;
      model log_zp = sickleYN posmalaria / solution;
      format sickleYN sickleYN. posmalaria yesno.;
run;

*-------------;
* Recent fever;
*-------------;

*Fever in 24 h: Test for interaction and confounding between sickleYN and
fever;
proc surveyreg data=samplesize;
      class sickleYN fever24;
      model log_zp = sickleYN fever24 sickleYN*fever24 / solution;
      cluster cluster;
      format sickleYN sickleYN. fever24 yesno.;
run;

*confounding: method 1;
proc surveyreg data=samplesize;
      class sickleYN fever24;
      cluster cluster;
      model log_zp = sickleYN fever24 / solution;
      lsmeans sickleYN / diff;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN fever24;
      cluster cluster;
      model log_zp = sickleYN fever24 / solution;
      format sickleYN sickleYN. fever24 yesno.;
run;

*-----;
* Thal ;
*-----;
proc freq data=samplesize;
      tables thalYN;
run;
*Interaction: Test for interaction and confounding between sickleYN and
thalassemia;
proc surveyreg data=samplesize;
      class sickleYN;
      cluster cluster;
      model log_zp = sickleYN thalYN sickleYN*thalYN / solution;
      format sickleYN sickleYN. thalYN yesno.;
run;
```

```
*confounding: method 2 (betas);
*SLR for crude beta;
proc surveyreg data=samplesize;
      class sickleYN;
      cluster cluster;
      model log_zp = sickleYN / solution;
      format sickleYN sickleYN.;
run;

*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN;
      cluster cluster;
      model log_zp = sickleYN thalYN / solution;
      format sickleYN sickleYN. thalYN yesno.;
run;

*-----;
*G6PD ;
*-----;

*Interaction: Test for interaction and confounding between sickleYN and g6pd;
proc surveyreg data=samplesize;
      class sickleYN;
      cluster cluster;
      model log_zp = sickleYN g6pd_def sickleYN*g6pd_def / solution;
      format sickleYN sickleYN. g6pd_def yesno.;
run;

*confounding: method 2 (betas);
*SLR for crude beta;
proc surveyreg data=samplesize;
      class sickleYN;
      cluster cluster;
      model log_zp = sickleYN / solution;
      format sickleYN sickleYN.;
run;

*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN;
      cluster cluster;
      model log_zp = sickleYN g6pd_def / solution;
      format sickleYN sickleYN. g6pd_def yesno.;
run;

**************************;
*Check for collinearity   *;
**************************;

proc corr data=samplesize;
      var log_zp sickleYN sprinkles24 posmalaria cf stunted education tea24
rbp thalYN g6pd_def;
      with sickleYN sprinkles24 posmalaria cf stunted education tea24 rbp
```

```
thalYN g6pd_def;
run;

*----------------------------;
* Exhaustive model          ;
*----------------------------;

*include all terms that were found to be significant for interaction or
confounding;
*interaction: sickleYN*sprinkles24 sickleYN*posmalaria
*sig in interaction model: inflammation1 rbp stunted fever24
*confounding: inflammation1 stunted education tea24 thalYN g6pd_def;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = sickleYN sprinkles24 sickleYN*sprinkles24 posmalaria
sickleYN*posmalaria
            inflammation1 rbp stunted fever24 education tea24 thalYN
g6pd_def/ solution;
      format sickleYN sickleYN. sprinkles24 yesno. posmalaria yesno.
inflammation1 yesno. stunted stunted.
      fever24 yesno. education yesno. tea24 yesno. thalYN yesno. g6pd_def
yesno.;
run;

*Add SF and TfR;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = sickleYN sprinkles24 sickleYN*sprinkles24 posmalaria
sickleYN*posmalaria
            inflammation1 rbp stunted fever24 education tea24 thalYN g6pd_def
log_sf log_tfr/ solution;
      format sickleYN sickleYN. sprinkles24 yesno. posmalaria yesno.
inflammation1 yesno. stunted stunted.
      fever24 yesno. education yesno. tea24 yesno. thalYN yesno. g6pd_def
yesno.;
run;

*----------------------------;
* Reduced model             ;
*----------------------------;

*without education;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = sickleYN sprinkles24 sickleYN*sprinkles24 posmalaria
sickleYN*posmalaria
            inflammation1 rbp stunted fever24 tea24 thalYN g6pd_def/
solution;
      format sickleYN sickleYN. sprinkles24 yesno. posmalaria yesno.
inflammation1 yesno. stunted stunted.
      fever24 yesno. tea24 yesno. thalYN yesno. g6pd_def yesno.;
run;

*without thal;
```

```
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = sickleYN sprinkles24 sickleYN*sprinkles24 posmalaria
sickleYN*posmalaria
            inflammation1 rbp stunted fever24 education tea24 g6pd_def/
solution;
      format sickleYN sickleYN. sprinkles24 yesno. posmalaria yesno.
inflammation1 yesno. stunted stunted.
      fever24 yesno. education yesno. tea24 yesno. g6pd_def yesno.;
run;

*without sprinkles;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = sickleYN sickleYN*sprinkles24 posmalaria
sickleYN*posmalaria
            inflammation1 rbp stunted fever24 education tea24 thalYN
g6pd_def/ solution;
      format sickleYN sickleYN. posmalaria yesno. inflammation1 yesno.
stunted stunted.
      fever24 yesno. education yesno. tea24 yesno. thalYN yesno. g6pd_def
yesno.;
run;

*without g6pd;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = sickleYN sprinkles24 sickleYN*sprinkles24 posmalaria
sickleYN*posmalaria
            inflammation1 rbp stunted fever24 education tea24 thalYN /
solution;
      format sickleYN sickleYN. sprinkles24 yesno. posmalaria yesno.
inflammation1 yesno. stunted stunted.
      fever24 yesno. education yesno. tea24 yesno. thalYN yesno.;
run;

*without fever;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = sickleYN sprinkles24 sickleYN*sprinkles24 posmalaria
sickleYN*posmalaria
            inflammation1 rbp stunted education tea24 thalYN g6pd_def/
solution;
      format sickleYN sickleYN. sprinkles24 yesno. posmalaria yesno.
inflammation1 yesno. stunted stunted.
      education yesno. tea24 yesno. thalYN yesno. g6pd_def yesno.;
run;

*without tea;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = sickleYN sprinkles24 sickleYN*sprinkles24 posmalaria
sickleYN*posmalaria
            inflammation1 rbp stunted fever24 education thalYN g6pd_def/
```

```
solution;
      format sickleYN sickleYN. sprinkles24 yesno. posmalaria yesno.
inflammation1 yesno. stunted stunted.
      fever24 yesno. education yesno. thalYN yesno. g6pd_def yesno.;
run;

*without RBP;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = sickleYN sprinkles24 sickleYN*sprinkles24 posmalaria
sickleYN*posmalaria
            inflammation1 stunted fever24 education tea24 thalYN g6pd_def/
solution;
      format sickleYN sickleYN. sprinkles24 yesno. posmalaria yesno.
inflammation1 yesno. stunted stunted.
      fever24 yesno. education yesno. tea24 yesno. thalYN yesno. g6pd_def
yesno.;
run;

*without stunted;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = sickleYN sprinkles24 sickleYN*sprinkles24 posmalaria
sickleYN*posmalaria
            inflammation1 rbp fever24 education tea24 thalYN g6pd_def/
solution;
      format sickleYN sickleYN. sprinkles24 yesno. posmalaria yesno.
inflammation1 yesno. stunted stunted.
      fever24 yesno. education yesno. tea24 yesno. thalYN yesno. g6pd_def
yesno.;
run;

****Reduced model****;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = sickleYN  posmalaria sickleYN*posmalaria inflammation1 /
solution;
      format sickleYN sickleYN. posmalaria yesno. inflammation1 yesno.;
run;

****ADD back hemoglobinopathies and thal as 3-levels****;
proc surveyreg data=samplesize;
      class thal;
      cluster cluster;
      model log_zp = sickleYN  posmalaria sickleYN*posmalaria inflammation1
g6pd_def thal/ solution;
      format sickleYN sickleYN. posmalaria yesno. inflammation1 yesno.
g6pd_def yesno. thal thal.;
run;

****ADD other iron indicators into model to see if R2 becomes larger****;
proc surveyreg data=samplesize;
      class thal;
      cluster cluster;
```

```
      model log_zp = sickleYN  posmalaria sickleYN*posmalaria inflammation1
g6pd_def thal log_sf log_tfr / solution;
      format sickleYN sickleYN. posmalaria yesno. inflammation1 yesno.
g6pd_def yesno. thal thal.;
run;

************REDUCED MODEL WITH SF AND TFR***************;
*EXHAUSTIVE MODEL*;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = sickleYN sprinkles24 sickleYN*sprinkles24 posmalaria
sickleYN*posmalaria
            inflammation1 rbp stunted fever24 education tea24 thalYN g6pd_def
log_sf log_tfr/ solution;
      format sickleYN sickleYN. sprinkles24 yesno. posmalaria yesno.
inflammation1 yesno. stunted stunted.
      fever24 yesno. education yesno. tea24 yesno. thalYN yesno. g6pd_def
yesno.;
run;

*removed stunted;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = sickleYN sprinkles24 sickleYN*sprinkles24 posmalaria
sickleYN*posmalaria
            inflammation1 rbp fever24 education tea24 thalYN g6pd_def log_sf
log_tfr/ solution;
      format sickleYN sickleYN. sprinkles24 yesno. posmalaria yesno.
inflammation1 yesno.
      fever24 yesno. education yesno. tea24 yesno. thalYN yesno. g6pd_def
yesno.;
run;

*removed recent tea;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = sickleYN sprinkles24 sickleYN*sprinkles24 posmalaria
sickleYN*posmalaria
            inflammation1 rbp stunted fever24 education thalYN g6pd_def
log_sf log_tfr/ solution;
      format sickleYN sickleYN. sprinkles24 yesno. posmalaria yesno.
inflammation1 yesno. stunted stunted.
      fever24 yesno. education yesno. thalYN yesno. g6pd_def yesno.;
run;

*removed sprinkles;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = sickleYN sickleYN*sprinkles24 posmalaria
sickleYN*posmalaria
            inflammation1 rbp stunted fever24 education tea24 thalYN g6pd_def
log_sf log_tfr/ solution;
      format sickleYN sickleYN. posmalaria yesno. inflammation1 yesno.
stunted stunted.
```

```
      fever24 yesno. education yesno. tea24 yesno. thalYN yesno. g6pd_def
yesno.;
run;

*removed thal;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = sickleYN sprinkles24 sickleYN*sprinkles24 posmalaria
sickleYN*posmalaria
            inflammation1 rbp stunted fever24 education tea24 g6pd_def log_sf
log_tfr/ solution;
      format sickleYN sickleYN. sprinkles24 yesno. posmalaria yesno.
inflammation1 yesno. stunted stunted.
      fever24 yesno. education yesno. tea24 yesno. g6pd_def yesno.;
run;

*removed mat educ;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = sickleYN sprinkles24 sickleYN*sprinkles24 posmalaria
sickleYN*posmalaria
            inflammation1 rbp stunted fever24 tea24 thalYN g6pd_def log_sf
log_tfr/ solution;
      format sickleYN sickleYN. sprinkles24 yesno. posmalaria yesno.
inflammation1 yesno. stunted stunted.
      fever24 yesno. tea24 yesno. thalYN yesno. g6pd_def yesno.;
run;

*removed g6pd;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = sickleYN sprinkles24 sickleYN*sprinkles24 posmalaria
sickleYN*posmalaria
            inflammation1 rbp stunted fever24 education tea24 thalYN log_sf
log_tfr/ solution;
      format sickleYN sickleYN. sprinkles24 yesno. posmalaria yesno.
inflammation1 yesno. stunted stunted.
      fever24 yesno. education yesno. tea24 yesno. thalYN yesno.;
run;

*removed recent fever;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = sickleYN sprinkles24 sickleYN*sprinkles24 posmalaria
sickleYN*posmalaria
            inflammation1 rbp stunted education tea24 thalYN g6pd_def log_sf
log_tfr/ solution;
      format sickleYN sickleYN. sprinkles24 yesno. posmalaria yesno.
inflammation1 yesno. stunted stunted.
      education yesno. tea24 yesno. thalYN yesno. g6pd_def yesno.;
run;

***FINAL MODEL WITH SF AND TFR****;
proc surveyreg data=samplesize;
```

```
      class thal;
      cluster cluster;
      model log_zp = sickleYN posmalaria inflammation1 rbp thal g6pd_def
log_sf log_tfr/ solution;
      format sickleYN sickleYN. posmalaria yesno. inflammation1 yesno. thal
thal. g6pd_def yesno.;
run;
proc surveyreg data=samplesize;
      class thal sickle_cl;
      cluster cluster;
      model log_zp = sickle_cl posmalaria inflammation1 rbp thal g6pd_def
log_sf log_tfr/ solution;
      format sickle_cl sickle. posmalaria yesno. inflammation1 yesno. thal
thal. g6pd_def yesno.;
run;

*--------------------------;
* LSMEANS for interaction   ;
*--------------------------;

*Sickle*malaria;

**Using Proc means**;
proc sort data=samplesize;
      by sickleYN posmalaria;
run;
proc means data=samplesize n mean clm;
      var log_zp;
      by sickleYN posmalaria;
run;


**Using proc surveymeans to account for cluster design**;
proc surveymeans data=samplesize;
   class thal;
   cluster cluster;
   domain sickleYN*posmalaria;
   var log_zp;
   format sickleYN sickleYN. posmalaria yesno.;
run;
```

3. 'thesis_tfr_sickle_aug19'

This file contains regression analysis for the relationship between TfR and sickle cell.

```
*******************************;
* Kiersten Derby                *;
* Thesis Regression Analysis  *;
* Feb 6, 2012                    *;
*******************************;

*-------------------------------------------------------;
```

```
*Call in 'samplesize' data from thesis_datasteps.sas    ;
*-------------------------------------------------------;

%include "H:\Thesis\thesis_datasteps.sas";

*Set reference values;
proc format;
      value sickleYN
            0 = "1 No Sickle"
            1 = "0 Yes Sickle"
      ; *No sickle as reference;

      value cf
            1 = "4 Reference"
            2 = "1 Incubation"
            3 = "2 Early Convalescence"
            4 = "3 Late Convalescence"
      ; *Reference as reference;

      value stunted
            0 = "1 Not stunted"
            1 = "0 Stunted"
      ; *No stunting as reference;

      value wasted
            0 = "1 Not wasted"
            1 = "0 Wasted"
      ; *No wasting as reference;

      value underwt
            0 = "1 Not underweight"
            1 = "0 Underweight"
      ; *Not underweight as reference;

      value poor
            0 = "1 Not poor"
            1 = "0 Poor"
      ; *Not poor as reference;

      value yesno
            0 = "2 No"
            1 = "1 Yes"
            . = "0 Missing"
      ; *NO as reference;

      value sickle
            1 = "3 Normal"
            2 = "2 Trait"
            3 = "1 Disease"
            . = "0 Missing"
      ; *normal as reference;

run;
*-------------------------------------------------------------;
```

```
*Regression Analysis: check for interaction and confounding   ;
*-------------------------------------------------------------;
**********************;
* TfR and Sickle      *;
**********************;

*Check for linearity for continuous variables age and RBP with TfR;
title;
ods listing;
symbol2 interpol=rl value=star color=blue;
proc gplot data=samplesize;
      plot log_tfr*rbp log_tfr*childagemonths_lab;
run;

*-----;
* AGE ;
*-----;
*Interaction: Test for interaction and confounding between sickleYN and age;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = sickleYN childagemonths_lab sickleYN*childagemonths_lab
/ solution;
      format sickleYN sickleYN.;
run;

*confounding: method 2 (betas);
*SLR for crude beta;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = sickleYN / solution;
      format sickleYN sickleYN.;
run;

*MLR for adjusted beta;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = sickleYN childagemonths_lab / solution;
      format sickleYN sickleYN.;
run;


*-----;
* SEX ;
*-----;
*Interaction: Test for interaction and confounding between sickleYN and sex;
proc surveyreg data=samplesize;
      model log_tfr = sickleYN sex1 sickleYN*sex1 / solution;
      cluster cluster;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
*reference sex is 1 (female);
```

```
      cluster cluster;
      model log_tfr = sickleYN sex1 / solution;
      format sickleYN sickleYN.;
run;


*---------;
* Inflamm ; *changed to include any inflammation versus no inflammation;
*---------;

*inflammation vs no inflammation;
proc surveyreg data=samplesize;
      model log_tfr = sickleYN inflammation1 sickleYN*inflammation1 /
solution;
      cluster cluster;
      format sickleYN sickleYN. inflammation1 yesno.;
run;

*confounding: method 2 (betas);

*MLR for adjusted beta;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = sickleYN inflammation1 / solution;
      format sickleYN sickleYN. inflammation1 yesno.;
run;

*----;
*RBP ;
*----;
*Interaction: Test for interaction and confounding between sickleYN and RBP;
proc surveyreg data=samplesize;
      model log_tfr = sickleYN rbp sickleYN*rbp / solution;
      cluster cluster;
      format sickleYN sickleYN.;
run;

*confounding: method 2 (betas);

*MLR for adjusted beta;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = sickleYN rbp / solution;
      format sickleYN sickleYN.;
run;


*---------;
*stunting ;
*---------;
*Interaction: Test for interaction and confounding between sickleYN and
stunting;
proc surveyreg data=samplesize;
      model log_tfr = sickleYN stunted sickleYN*stunted / solution;
```

```
      cluster cluster;
      format sickleYN sickleYN. stunted stunted.;
run;

*confounding: method 2 (betas);

*MLR for adjusted beta;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = sickleYN stunted / solution;
      format sickleYN sickleYN. stunted stunted.;
run;


*---------;
*Wasting  ;
*---------;
*Interaction: Test for interaction and confounding between sickleYN and
wasting;
proc surveyreg data=samplesize;
      model log_tfr = sickleYN wasted sickleYN*wasted / solution;
      cluster cluster;
      format sickleYN sickleYN. wasted wasted.;
run;

*confounding: method 2 (betas);

*MLR for adjusted beta;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = sickleYN wasted / solution;
      format sickleYN sickleYN. wasted wasted.;
run;

*-------------;
*Underweight  ;
*-------------;
*Interaction: Test for interaction and confounding between sickleYN and
underweight;
proc surveyreg data=samplesize;
      model log_tfr = sickleYN underwt sickleYN*underwt / solution;
      cluster cluster;
      format sickleYN sickleYN. underwt underwt.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = sickleYN underwt / solution;
      format sickleYN sickleYN. underwt underwt.;
run;

*-------------;
```

```
* SES         ;
*-------------;
*SES: Test for interaction and confounding between sickleYN and SES where 1=
quintiles 1 or 2, 0 = quintiles 3-5;
*Interaction;
proc surveyreg data=samplesize;
      class sickleYN poor;
      model log_tfr = sickleYN poor sickleYN*poor / solution;
      cluster cluster;
      format sickleYN sickleYN. poor poor.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN poor;
      cluster cluster;
      model log_tfr = sickleYN poor / solution;
      format sickleYN sickleYN. poor poor.;
run;

*-------------;
* maternal edu;
*-------------;

*Maternal Education: Test for interaction and confounding between sickleYN
and education;
*Interaction;
proc surveyreg data=samplesize;
      model log_tfr = sickleYN education sickleYN*education / solution;
      cluster cluster;
      format sickleYN sickleYN. education yesno.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = sickleYN education / solution;
      format sickleYN sickleYN. education yesno.;
run;

*-------------;
* recent tea  ;
*-------------;

*Interaction: Test for interaction and confounding between sickleYN and tea;
proc surveyreg data=samplesize;
      class sickleYN tea24;
      model log_tfr = sickleYN tea24 sickleYN*tea24 / solution;
      cluster cluster;
      format sickleYN sickleYN. tea24 yesno.;
run;
```

```
*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN tea24;
      cluster cluster;
      model log_tfr = sickleYN tea24 / solution;
      format sickleYN sickleYN. tea24 yesno.;
run;

*-------------;
* Sprinkles;
*-------------;
*Interaction: Test for interaction and confounding between sickleYN and
sprinkles;
proc surveyreg data=samplesize;
      class sickleYN sprinkles24;
      model log_tfr = sickleYN sprinkles24 sickleYN*sprinkles24 / solution;
      cluster cluster;
      format sickleYN sickleYN. sprinkles24 yesno.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN sprinkles24;
      cluster cluster;
      model log_tfr = sickleYN sprinkles24 / solution;
      format sickleYN sickleYN. sprinkles24 yesno.;
run;

*-------------;
* Malaria     ;
*-------------;

*Interaction: Test for interaction and confounding between sickleYN and
malaria;
proc surveyreg data=samplesize;
      class sickleYN posmalaria;
      model log_tfr = sickleYN posmalaria sickleYN*posmalaria / solution;
      cluster cluster;
      format sickleYN sickleYN. posmalaria yesno.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN posmalaria;
      cluster cluster;
      model log_tfr = sickleYN posmalaria / solution;
      format sickleYN sickleYN. posmalaria yesno.;
run;

*-------------;
* Recent fever;
```

```
*-------------;
*Fever in 24 h: Test for interaction and confounding between sickleYN and
fever;
proc surveyreg data=samplesize;
      class sickleYN fever24;
      model log_tfr = sickleYN fever24 sickleYN*fever24 / solution;
      cluster cluster;
      format sickleYN sickleYN. fever24 yesno.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN fever24;
      cluster cluster;
      model log_tfr = sickleYN fever24 / solution;
      format sickleYN sickleYN. fever24 yesno.;
run;

*-----;
* Thal ;
*-----;
*Interaction: Test for interaction and confounding between sickleYN and
thalassemia;
proc surveyreg data=samplesize;
      class sickleYN;
      cluster cluster;
      model log_tfr = sickleYN thalYN sickleYN*thalYN / solution;
      format sickleYN sickleYN. thalYN yesno.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN;
      cluster cluster;
      model log_tfr = sickleYN thalYN / solution;
      format sickleYN sickleYN. thalYN yesno.;
run;

*-----;
*G6PD ;
*-----;
*Interaction: Test for interaction and confounding between sickleYN and g6pd;
proc surveyreg data=samplesize;
      class sickleYN;
      cluster cluster;
      model log_tfr = sickleYN g6pd_def sickleYN*g6pd_def / solution;
      format sickleYN sickleYN. g6pd_def yesno.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
```

```
      class sickleYN;
      cluster cluster;
      model log_tfr = sickleYN g6pd_def / solution;
      format sickleYN sickleYN. g6pd_def yesno.;
run;


*--------------------------;
* Exhaustive model         ;
*--------------------------;

******* inflammation1 instead of cf ********;
*interaction terms = sprinkles*sickle, malaria*sickle . Include rbp,
stunting,
      underweight, recent fever because sig along in interaction test;
*confounders = stunting, education, malaria, g6pd;
proc surveyreg data=samplesize;
/*    class sickleYN sprinkles24 posmalaria inflammation1 stunted underwt
fever24 education g6pd_def;*/
      cluster cluster;
      model log_tfr = sickleYN inflammation1 sprinkles24 sickleYN*sprinkles24
posmalaria sickleYN*posmalaria
            rbp stunted underwt fever24 education g6pd_def/ solution;
      format sickleYN sickleYN. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.;
run;

**exhaustive model with other iron indicators**;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = sickleYN inflammation1 sprinkles24 sickleYN*sprinkles24
posmalaria sickleYN*posmalaria
            rbp stunted underwt fever24 education g6pd_def log_zp log_sf /
solution;
      format sickleYN sickleYN. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.;
run;

*--------------------------;
* Reduced model            ;
*--------------------------;

*take out education;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = sickleYN inflammation1 sprinkles24 sickleYN*sprinkles24
posmalaria sickleYN*posmalaria
            rbp stunted underwt fever24 g6pd_def/ solution;
      format sickleYN sickleYN. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno. stunted stunted.
            g6pd_def yesno. underwt underwt. fever24 yesno.;
run;
```

```
*take out underweight;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = sickleYN inflammation1 sprinkles24 sickleYN*sprinkles24
posmalaria sickleYN*posmalaria
            rbp stunted fever24 education g6pd_def/ solution;
      format sickleYN sickleYN. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. fever24 yesno.;
run;

*Take out G6PD;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = sickleYN inflammation1 sprinkles24 sickleYN*sprinkles24
posmalaria sickleYN*posmalaria
            rbp stunted underwt fever24 education/ solution;
      format sickleYN sickleYN. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno. stunted stunted.
            education yesno. underwt underwt. fever24 yesno.;
run;

*Take out recent fever;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = sickleYN inflammation1 sprinkles24 sickleYN*sprinkles24
posmalaria sickleYN*posmalaria
            rbp stunted underwt education g6pd_def/ solution;
      format sickleYN sickleYN. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt.;
run;

*Take out RBP;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = sickleYN inflammation1 sprinkles24 sickleYN*sprinkles24
posmalaria sickleYN*posmalaria
            stunted underwt fever24 education g6pd_def/ solution;
      format sickleYN sickleYN. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.;
run;

*Take out Sprinkles;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = sickleYN inflammation1 sickleYN*sprinkles24 posmalaria
sickleYN*posmalaria
            rbp stunted underwt fever24 education g6pd_def/ solution;
      format sickleYN sickleYN. inflammation1 yesno. posmalaria yesno.
stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.;
```

```
run;


*****REDUCED MODEL WITH OTHER IRON INDICATORS INCLUDED*****;
**exhaustive model with other iron indicators**;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = sickleYN inflammation1 sickleYN*sprinkles24 sprinkles24
sickleYN*posmalaria posmalaria
            rbp stunted underwt fever24 education g6pd_def log_zp log_sf /
solution;
      format sickleYN sickleYN. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.;
run;

*remove underwt;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = sickleYN inflammation1 sickleYN*sprinkles24 sprinkles24
sickleYN*posmalaria posmalaria
            rbp stunted fever24 education g6pd_def log_zp log_sf / solution;
      format sickleYN sickleYN. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.;
run;

*remove fever;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = sickleYN inflammation1 sickleYN*sprinkles24 sprinkles24
sickleYN*posmalaria posmalaria
            rbp stunted underwt education g6pd_def log_zp log_sf / solution;
      format sickleYN sickleYN. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.;
run;

*remove educ;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = sickleYN inflammation1 sickleYN*sprinkles24 sprinkles24
sickleYN*posmalaria posmalaria
            rbp stunted underwt fever24 g6pd_def log_zp log_sf / solution;
      format sickleYN sickleYN. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.;
run;

*remove sprinkles;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = sickleYN inflammation1 sickleYN*sprinkles24 posmalaria
sickleYN*posmalaria
```

```
            rbp stunted underwt fever24 education g6pd_def log_zp log_sf /
solution;
      format sickleYN sickleYN. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.;
run;

*---------------------------;
* Final   model             ;
*---------------------------;
**Also drop interaction terms (because p>0.01);

proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = sickleYN inflammation1 posmalaria stunted / solution;
      format sickleYN sickleYN. inflammation1 yesno. posmalaria yesno.
stunted stunted.;
run;

**add back in thal and G6PD def. THIS IS THE FINAL MODEL**;
proc surveyreg data=samplesize;
      class thal;
      cluster cluster;
      model log_tfr = sickleYN inflammation1 posmalaria stunted g6pd_def
thal/ solution;
      format sickleYN sickleYN. inflammation1 yesno. posmalaria yesno.
stunted stunted.
            g6pd_def yesno. thal thal.;
run;

**add in other iron indicators to look at change in R2 (not index because TfR
used to define index**;
proc surveyreg data=samplesize;
      class thal;
      cluster cluster;
      model log_tfr = sickleYN inflammation1 posmalaria stunted g6pd_def thal
            log_sf log_zp / solution;
      format sickleYN sickleYN. inflammation1 yesno. posmalaria yesno.
stunted stunted.
            g6pd_def yesno. thal thal.;
run;


****FINAL MODEL WITH OTHER IRON INDICATORS****;
proc surveyreg data=samplesize;
      class thal;
      cluster cluster;
      model log_tfr = sickleYN inflammation1 posmalaria rbp stunted thal
g6pd_def log_zp log_sf / solution;
      format sickleYN sickleYN. thal thal. inflammation1 yesno. sprinkles24
yesno. posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.;
run;
```

```
*look for concentration effect of sickle;
proc surveyreg data=samplesize;
      class thal sickle_cl;
      cluster cluster;
      model log_tfr = sickle_cl inflammation1 posmalaria rbp stunted thal
g6pd_def log_zp log_sf / solution;
      format sickle_cl sickle. thal thal. inflammation1 yesno. sprinkles24
yesno. posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.;
run;

**************************;
*Check for collinearity   *;
**************************;

proc corr data=samplesize;
      var log_tfr sickleYN cf sprinkles24 posmalaria stunted education;
      with sickleYN cf sprinkles24 posmalaria stunted education;
run;

proc reg data=work.samplesize plots(unpack);
      model log_tfr = sickleYN cf sprinkles24 posmalaria stunted
int_inc_sickle
            int_sprinkles_sickle int_posmalaria_sickle / partial vif;
run; quit;

*******************************************;
** Look at diagnostics of associative model **;
*******************************************;

ods graphics on;
ods exclude rfplot where= (_label_?'Intercept');

proc reg data=work.samplesize plots(unpack);
      model log_tfr = sickleYN cf sprinkles24 posmalaria stunted
int_inc_sickle
            int_sprinkles_sickle int_posmalaria_sickle / partial vif;
      output out=work.regdata2 R=resid P=yhat rstudent=jackknife cookd=cooksD
H=leverage;
run; quit;

proc print data=work.regdata2;
run;

*check normality on residuals;
proc univariate data=work.regdata2;
      var resid;
run;

**determine critical values;
data work.tcrit;
      tcrit = tinv(0.975,804); *n-k-2 = 814-8-2 = 804;
run;
proc print data=work.tcrit;
```

```
run;
**crit value = 1.96292;

*leverage cut-off = 2(k+1)/n = 2(8+1)/804 = 0.022388;

*check for outliers and influential values;
proc print data=work.regdata2;
      where abs(jackknife) > 1.96292 or cooksD > 1 or leverage >0.022388;
run;

title 'jackknife';
proc print data=work.regdata2;
      where abs(jackknife) > 1.96292;
run;

title 'cooksD';
proc print data=work.regdata2;
      where cooksD > 1;
run;

title 'leverage';
proc print data=work.regdata2;
      where leverage >0.022388;
run;
```

4. 'thesis_sf_sickle_aug19'

This file contains code for the regression analysis of the relationship between SF and sickle cell. Also contains code for mean SF by sickle cell and poverty level (interaction term).

```
*********************************;
* Kiersten Derby                 *;
* Thesis Regression Analysis  *;
* Feb 6, 2012                              *;
*********************************;

*-------------------------------------------------------------;
*Call in 'samplesize' data from thesis_datasteps.sas    ;
*-------------------------------------------------------------;

%include "H:\Thesis\thesis_datasteps_aug19.sas";

*-------------------------------------------------------------------;
*Regression Analysis: check for interaction and confounding   ;
*-------------------------------------------------------------------;
***********************;
* Ferritin and Sickle *;
***********************;

*Check for linearity for continuous variables age and RBP with TfR;
title;
ods listing;
```

```
symbol2 interpol=rl value=star color=blue;
proc gplot data=samplesize;
      plot log_sf*rbp log_sf*childagemonths_lab;
run;

*-----;
* AGE ;
*-----;
*Interaction: Test for interaction and confounding between sickleYN and age;
proc surveyreg data=samplesize;
      class sickleYN;
      cluster cluster;
      model log_sf = sickleYN childagemonths_lab sickleYN*childagemonths_lab
/ solution;
      format sickleYN sickleYN.;
run;

*confounding: method 2 (betas);
*SLR for crude beta;
proc surveyreg data=samplesize;
      class sickleYN;
      cluster cluster;
      model log_sf = sickleYN / solution;
      format sickleYN sickleYN.;
run;

*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN;
      cluster cluster;
      model log_sf = sickleYN childagemonths_lab / solution;
      format sickleYN sickleYN.;
run;


*-----;
* SEX ;
*-----;
*Interaction: Test for interaction and confounding between sickleYN and sex;
proc surveyreg data=samplesize;
      class sickleYN sex1;
      model log_sf = sickleYN sex1 sickleYN*sex1 / solution;
      cluster cluster;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN sex1; *reference sex is 1 (female);
      cluster cluster;
      model log_sf = sickleYN sex1 / solution;
      format sickleYN sickleYN.;
run;
```

```
*---------;
* Inflamm ;
*---------;
*Inflammation: Test for interaction and confounding between sickleYN and
inflammation;
proc surveyreg data=samplesize;
      model log_sf = sickleYN inflammation1 sickleYN*inflammation1 /
solution;
      cluster cluster;
      format sickleYN sickleYN. inflammation1 yesno.;
run;

*confounding: method 2 (betas);
proc surveyreg data=samplesize;
      cluster cluster;
      model log_sf = sickleYN inflammation1 / solution;
      format sickleYN sickleYN. inflammation1 yesno.;
run;

*----;
*RBP ;
*----;
*Interaction: Test for interaction and confounding between sickleYN and RBP;
proc surveyreg data=samplesize;
      class sickleYN;
      model log_sf = sickleYN rbp sickleYN*rbp / solution;
      cluster cluster;
      format sickleYN sickleYN.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN;
      cluster cluster;
      model log_sf = sickleYN rbp / solution;
      format sickleYN sickleYN.;
run;


*---------;
*stunting ;
*---------;
*Interaction: Test for interaction and confounding between sickleYN and
stunting;
proc surveyreg data=samplesize;
      class sickleYN stunted;
      model log_sf = sickleYN stunted sickleYN*stunted / solution;
      cluster cluster;
      format sickleYN sickleYN. stunted stunted.;
run;

*confounding: method 2 (betas);
```

```
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN stunted;
      cluster cluster;
      model log_sf = sickleYN stunted / solution;
      format sickleYN sickleYN. stunted stunted.;
run;


*---------;
*Wasting  ;
*---------;
*Interaction: Test for interaction and confounding between sickleYN and
wasting;
proc surveyreg data=samplesize;
      class sickleYN wasted;
      model log_sf = sickleYN wasted sickleYN*wasted / solution;
      cluster cluster;
      format sickleYN sickleYN. wasted wasted.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN wasted;
      cluster cluster;
      model log_sf = sickleYN wasted / solution;
      format sickleYN sickleYN. wasted wasted.;
run;

*-------------;
*Underweight  ;
*-------------;
*Interaction: Test for interaction and confounding between sickleYN and
underwt;
proc surveyreg data=samplesize;
      class sickleYN underwt;
      model log_sf = sickleYN underwt sickleYN*underwt / solution;
      cluster cluster;
      format sickleYN sickleYN. underwt underwt.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN underwt;
      cluster cluster;
      model log_sf = sickleYN underwt / solution;
      format sickleYN sickleYN. underwt underwt.;
run;

*-------------;
* SES         ;
*-------------;
```

```
*SES: Test for interaction and confounding between sickleYN and SES where 1=
quintiles 1 or 2, 0 = quintiles 3-5;
*Interaction;
proc surveyreg data=samplesize;
      class sickleYN poor;
      model log_sf = sickleYN poor sickleYN*poor / solution;
      cluster cluster;
      format sickleYN sickleYN. poor poor.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN poor;
      cluster cluster;
      model log_sf = sickleYN poor / solution;
      format sickleYN sickleYN. poor poor.;
run;

*-------------;
* maternal edu;
*-------------;

*Maternal Education: Test for interaction and confounding between sickleYN
and education;
*Interaction;
proc surveyreg data=samplesize;
      model log_sf = sickleYN education sickleYN*education / solution;
      cluster cluster;
      format sickleYN sickleYN. education yesno.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_sf = sickleYN education / solution;
      format sickleYN sickleYN. education yesno.;
run;

*-------------;
* recent tea  ;
*-------------;

*Interaction: Test for interaction and confounding between sickleYN and tea;
proc surveyreg data=samplesize;
      class sickleYN tea24;
      model log_sf = sickleYN tea24 sickleYN*tea24 / solution;
      cluster cluster;
      format sickleYN sickleYN. tea24 yesno.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
```

```
proc surveyreg data=samplesize;
      class sickleYN tea24;
      cluster cluster;
      model log_sf = sickleYN tea24 / solution;
      format sickleYN sickleYN. tea24 yesno.;
run;

*-------------;
* Sprinkles;
*-------------;

*Interaction: Test for interaction and confounding between sickleYN and
sprinkles;
proc surveyreg data=samplesize;
      class sickleYN sprinkles24;
      model log_sf = sickleYN sprinkles24 sickleYN*sprinkles24 / solution;
      cluster cluster;
      format sickleYN sickleYN. sprinkles24 yesno.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN sprinkles24;
      cluster cluster;
      model log_sf = sickleYN sprinkles24 / solution;
      format sickleYN sickleYN. sprinkles24 yesno.;
run;

*-------------;
* Malaria     ;
*-------------;
*Interaction: Test for interaction and confounding between sickleYN and
malaria;
proc surveyreg data=samplesize;
      class sickleYN posmalaria;
      model log_sf = sickleYN posmalaria sickleYN*posmalaria / solution;
      cluster cluster;
      format sickleYN sickleYN. posmalaria yesno.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN posmalaria;
      cluster cluster;
      model log_sf = sickleYN posmalaria / solution;
      format sickleYN sickleYN. posmalaria yesno.;
run;

*-------------;
* Recent fever;
*-------------;
```

```
*Fever in 24 h: Test for interaction and confounding between sickleYN and
fever;
proc surveyreg data=samplesize;
      class sickleYN fever24;
      model log_sf = sickleYN fever24 sickleYN*fever24 / solution;
      cluster cluster;
      format sickleYN sickleYN. fever24 yesno.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN fever24;
      cluster cluster;
      model log_sf = sickleYN fever24 / solution;
      format sickleYN sickleYN. fever24 yesno.;
run;


*-----;
* Thal ;
*-----;

*Interaction: Test for interaction and confounding between sickleYN and
thalassemia;
proc surveyreg data=samplesize;
      class sickleYN;
      cluster cluster;
      model log_sf = sickleYN thalYN sickleYN*thalYN / solution;
      format sickleYN sickleYN. thalYN yesno.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN;
      cluster cluster;
      model log_sf = sickleYN thalYN / solution;
      format sickleYN sickleYN. thalYN yesno.;
run;

*-----;
*G6PD ;
*-----;

*Interaction: Test for interaction and confounding between sickleYN and g6pd;
proc surveyreg data=samplesize;
      class sickleYN;
      cluster cluster;
      model log_sf = sickleYN g6pd_def sickleYN*g6pd_def / solution;
      format sickleYN sickleYN. g6pd_def yesno.;
run;

*confounding: method 2 (betas);
```

```
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN;
      cluster cluster;
      model log_sf = sickleYN g6pd_def / solution;
      format sickleYN sickleYN. g6pd_def yesno.;
run;


*----------------------------;
* Exhaustive model          ;
*----------------------------;

*include all terms that were found to be significant for interaction or
confounding;
*interaction terms = wasting*sickle underweight*sickle SES*sickle . Include
inflammation,age
      rbp, malaria parasitemia, recent fever because sig along in interaction
test;
*confounders = age, inflammation stage, rbp, wasting, underweight, ses,
education, tea, sprinkles, malaria, recent fever
      thal, G6PD;

proc surveyreg data=samplesize;
      cluster cluster;
      model log_sf = sickleYN wasted wasted*sickleYN underwt underwt*sickleYN
poor poor*sickleYN inflammation1 rbp
            posmalaria fever24 childagemonths_lab education tea24 sprinkles24
thalYN g6pd_def/ solution;
      format sickleYN sickleYN. wasted wasted. poor poor. inflammation1
yesno. sprinkles24 yesno. posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.
tea24 yesno. thalYN yesno. g6pd_def yesno.;
run;


***exhaustive model with other iron indicators;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_sf = sickleYN wasted wasted*sickleYN underwt underwt*sickleYN
poor poor*sickleYN inflammation1 rbp
            posmalaria fever24 childagemonths_lab education tea24 sprinkles24
thalYN g6pd_def log_zp log_tfr/ solution;
      format sickleYN sickleYN. wasted wasted. poor poor. inflammation1
yesno. sprinkles24 yesno. posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.
tea24 yesno. thalYN yesno. g6pd_def yesno.;
run;

*----------------------------;
* Reduced model             ;
*----------------------------;

*take out education;
```

```
proc surveyreg data=samplesize;
      cluster cluster;
      model log_sf = sickleYN wasted wasted*sickleYN underwt underwt*sickleYN
poor poor*sickleYN inflammation1 rbp
            posmalaria fever24 childagemonths_lab tea24 sprinkles24 thalYN
g6pd_def/ solution;
      format sickleYN sickleYN. wasted wasted. poor poor. inflammation1
yesno. sprinkles24 yesno. posmalaria yesno. stunted stunted.
            g6pd_def yesno. underwt underwt. fever24 yesno. tea24 yesno.
thalYN yesno. g6pd_def yesno.;
run;

*Take out RBP;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_sf = sickleYN wasted wasted*sickleYN underwt underwt*sickleYN
poor poor*sickleYN inflammation1
            posmalaria fever24 childagemonths_lab education tea24 sprinkles24
thalYN g6pd_def/ solution;
      format sickleYN sickleYN. wasted wasted. poor poor. inflammation1
yesno. sprinkles24 yesno. posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.
tea24 yesno. thalYN yesno. g6pd_def yesno.;
run;

*Take out G6PD def;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_sf = sickleYN wasted wasted*sickleYN underwt underwt*sickleYN
poor poor*sickleYN inflammation1 rbp
            posmalaria fever24 childagemonths_lab education tea24 sprinkles24
thalYN / solution;
      format sickleYN sickleYN. wasted wasted. poor poor. inflammation1
yesno. sprinkles24 yesno. posmalaria yesno. stunted stunted.
            education yesno. underwt underwt. fever24 yesno. tea24 yesno.
thalYN yesno. ;
run;

*Take out Sprinkles;
proc surveyreg data=samplesize;
      class sickleYN wasted underwt poor cf posmalaria fever24 education
tea24 g6pd_def;
      cluster cluster;
      model log_sf = sickleYN wasted wasted*sickleYN underwt underwt*sickleYN
poor poor*sickleYN cf rbp
            posmalaria fever24 childagemonths_lab education tea24 g6pd_def/
solution;
      format sickleYN sickleYN. wasted wasted. poor poor. cf cf. posmalaria
yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.
tea24 yesno. g6pd_def yesno.;
run;
*keep;
```

```
*Take out recent tea;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_sf = sickleYN wasted wasted*sickleYN underwt underwt*sickleYN
poor poor*sickleYN inflammation1 rbp
            posmalaria fever24 childagemonths_lab education sprinkles24
thalYN g6pd_def/ solution;
      format sickleYN sickleYN. wasted wasted. poor poor. inflammation1
yesno. sprinkles24 yesno. posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.
thalYN yesno. g6pd_def yesno.;
run;

*Take out SES;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_sf = sickleYN wasted wasted*sickleYN underwt underwt*sickleYN
poor*sickleYN inflammation1 rbp
            posmalaria fever24 childagemonths_lab education tea24 sprinkles24
thalYN g6pd_def/ solution;
      format sickleYN sickleYN. wasted wasted. inflammation1 yesno.
sprinkles24 yesno. posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.
tea24 yesno. thalYN yesno. g6pd_def yesno.;
run;

*Take out wasted;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_sf = sickleYN wasted*sickleYN underwt underwt*sickleYN poor
poor*sickleYN inflammation1 rbp
            posmalaria fever24 childagemonths_lab education tea24 sprinkles24
thalYN g6pd_def/ solution;
      format sickleYN sickleYN. poor poor. inflammation1 yesno. sprinkles24
yesno. posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.
tea24 yesno. thalYN yesno. g6pd_def yesno.;
run;

*Take out underweight;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_sf = sickleYN wasted wasted*sickleYN underwt*sickleYN poor
poor*sickleYN inflammation1 rbp
            posmalaria fever24 childagemonths_lab education tea24 sprinkles24
thalYN g6pd_def/ solution;
      format sickleYN sickleYN. wasted wasted. poor poor. inflammation1
yesno. sprinkles24 yesno. posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. fever24 yesno. tea24 yesno.
thalYN yesno. g6pd_def yesno.;
run;

*Take out age;
proc surveyreg data=samplesize;
```

```
      cluster cluster;
      model log_sf = sickleYN wasted wasted*sickleYN underwt underwt*sickleYN
poor poor*sickleYN inflammation1 rbp
            posmalaria fever24 education tea24 sprinkles24 thalYN g6pd_def/
solution;
      format sickleYN sickleYN. wasted wasted. poor poor. inflammation1
yesno. sprinkles24 yesno. posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.
tea24 yesno. thalYN yesno. g6pd_def yesno.;
run;

**REDUCED MODEL WITH OTHER IRON INDICATORS**;
***exhaustive model with other iron indicators;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_sf = sickleYN wasted wasted*sickleYN underwt underwt*sickleYN
poor poor*sickleYN inflammation1 rbp
            posmalaria fever24 childagemonths_lab education tea24 sprinkles24
thalYN g6pd_def log_zp log_tfr/ solution;
      format sickleYN sickleYN. wasted wasted. poor poor. inflammation1
yesno. sprinkles24 yesno. posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.
tea24 yesno. thalYN yesno. g6pd_def yesno.;
run;

*removed rbp;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_sf = sickleYN wasted wasted*sickleYN underwt underwt*sickleYN
poor poor*sickleYN inflammation1
            posmalaria fever24 childagemonths_lab education tea24 sprinkles24
thalYN g6pd_def log_zp log_tfr/ solution;
      format sickleYN sickleYN. wasted wasted. poor poor. inflammation1
yesno. sprinkles24 yesno. posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.
tea24 yesno. thalYN yesno. g6pd_def yesno.;
run;

*removed mat educ;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_sf = sickleYN wasted wasted*sickleYN underwt underwt*sickleYN
poor poor*sickleYN inflammation1 rbp
            posmalaria fever24 childagemonths_lab tea24 sprinkles24 thalYN
g6pd_def log_zp log_tfr/ solution;
      format sickleYN sickleYN. wasted wasted. poor poor. inflammation1
yesno. sprinkles24 yesno. posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.
tea24 yesno. thalYN yesno. g6pd_def yesno.;
run;

*removed tea;
proc surveyreg data=samplesize;
      cluster cluster;
```

```
      model log_sf = sickleYN wasted wasted*sickleYN underwt underwt*sickleYN
poor poor*sickleYN inflammation1 rbp
            posmalaria fever24 childagemonths_lab education sprinkles24
thalYN g6pd_def log_zp log_tfr/ solution;
      format sickleYN sickleYN. wasted wasted. poor poor. inflammation1
yesno. sprinkles24 yesno. posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.
tea24 yesno. thalYN yesno. g6pd_def yesno.;
run;

*removed sprinkles;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_sf = sickleYN wasted wasted*sickleYN underwt underwt*sickleYN
poor poor*sickleYN inflammation1 rbp
            posmalaria fever24 childagemonths_lab education tea24 thalYN
g6pd_def log_zp log_tfr/ solution;
      format sickleYN sickleYN. wasted wasted. poor poor. inflammation1
yesno. sprinkles24 yesno. posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.
tea24 yesno. thalYN yesno. g6pd_def yesno.;
run;

*removed age;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_sf = sickleYN wasted wasted*sickleYN underwt underwt*sickleYN
poor poor*sickleYN inflammation1 rbp
            posmalaria fever24 education tea24 thalYN g6pd_def log_zp
log_tfr/ solution;
      format sickleYN sickleYN. wasted wasted. poor poor. inflammation1
yesno. sprinkles24 yesno. posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.
tea24 yesno. thalYN yesno. g6pd_def yesno.;
run;

*removed wasted;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_sf = sickleYN wasted*sickleYN underwt underwt*sickleYN poor
poor*sickleYN inflammation1 rbp
            posmalaria fever24 childagemonths_lab education tea24 sprinkles24
thalYN g6pd_def log_zp log_tfr/ solution;
      format sickleYN sickleYN. wasted wasted. poor poor. inflammation1
yesno. sprinkles24 yesno. posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.
tea24 yesno. thalYN yesno. g6pd_def yesno.;
run;

*removed underwt;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_sf = sickleYN wasted wasted*sickleYN underwt*sickleYN poor
poor*sickleYN inflammation1 rbp
```

```
            posmalaria fever24 childagemonths_lab education tea24 sprinkles24
thalYN g6pd_def log_zp log_tfr/ solution;
      format sickleYN sickleYN. wasted wasted. poor poor. inflammation1
yesno. sprinkles24 yesno. posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.
tea24 yesno. thalYN yesno. g6pd_def yesno.;
run;


*----------------------------;
* Final   model             ;
*----------------------------;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_sf = sickleYN underwt poor poor*sickleYN inflammation1
            posmalaria fever24 sprinkles24 thalYN g6pd_def/ solution;
      format sickleYN sickleYN. poor poor. inflammation1 yesno. sprinkles24
yesno. posmalaria yesno. stunted stunted.
            g6pd_def yesno. underwt underwt. fever24 yesno. thalYN yesno.
g6pd_def yesno.;
run;

**add back hemoglobinopathies, thal 3 levels - THIS IS THE FINAL MODEL**;
proc surveyreg data=samplesize;
      class thal;
      cluster cluster;
      model log_sf = sickleYN underwt poor poor*sickleYN inflammation1
posmalaria fever24 sprinkles24 thal g6pd_def/ solution;
      format sickleYN sickleYN. poor poor. inflammation1 yesno. sprinkles24
yesno. posmalaria yesno. stunted stunted.
            g6pd_def yesno. underwt underwt. fever24 yesno. thal thal.
g6pd_def yesno.;
run;

**Add in other iron indicators to look at change in R-squared**;
proc surveyreg data=samplesize;
      class thal;
      cluster cluster;
      model log_sf = sickleYN underwt poor poor*sickleYN inflammation1
posmalaria fever24 sprinkles24 thal g6pd_def
            log_tfr log_zp / solution;
      format sickleYN sickleYN. poor poor. inflammation1 yesno. sprinkles24
yesno. posmalaria yesno. stunted stunted.
            g6pd_def yesno. underwt underwt. fever24 yesno. thal thal.
g6pd_def yesno.;
run;

***INTERACTION TERM _ MEANS***;
*Look at mean SF for interaction term Sickle*poor;

proc sort data=samplesize;
      by sickleYN poor;
run;
proc means data=samplesize n mean clm;
```

```
      var log_sf;
      by sickleYN poor;
run;

**Using proc surveymeans to account for cluster design**;
proc surveymeans data=samplesize;
   class thal;
   cluster cluster;
   domain sickleYN*poor;
   var log_sf;
   format sickleYN sickleYN. poor poor.;
run;

**FINAL MODEL WITH OTHER IRON INDICATORS **;

proc surveyreg data=samplesize;
      class thal;
      cluster cluster;
      model log_sf = sickleYN underwt poor poor*sickleYN inflammation1
            posmalaria fever24 thal g6pd_def log_zp log_tfr/ solution;
      format sickleYN sickleYN. wasted wasted. poor poor. inflammation1
yesno. sprinkles24 yesno. posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.
tea24 yesno. thal thal. g6pd_def yesno.;
run;

**************************;
*Check for collinearity   *;
**************************;

proc corr data=samplesize;
      var log_tfr sickleYN wasted underwt poor cf rbp posmalaria fever24
childagemonths_lab education sprinkles24
            thalYN g6pd_def;
      with sickleYN wasted underwt poor cf rbp posmalaria fever24
childagemonths_lab education sprinkles24 thalYN
            g6pd_def;
run;

***NEED TO DO THIS*****;
proc reg data=work.samplesize plots(unpack);
      model log_tfr = sickleYN cf sprinkles24 posmalaria stunted
int_inc_sickle
            int_sprinkles_sickle int_posmalaria_sickle / partial vif;
run; quit;

*******************************************;
** Look at diagnostics of associative model **;
*******************************************;

ods graphics on;
ods exclude rfplot where= (_label_?'Intercept');

proc reg data=work.samplesize plots(unpack);
```

```
      model log_tfr = sickleYN cf sprinkles24 posmalaria stunted
int_inc_sickle
            int_sprinkles_sickle int_posmalaria_sickle / partial vif;
      output out=work.regdata2 R=resid P=yhat rstudent=jackknife cookd=cooksD
H=leverage;
run; quit;

proc print data=work.regdata2;
run;

*check normality on residuals;
proc univariate data=work.regdata2;
      var resid;
run;

**determine critical values;
data work.tcrit;
      tcrit = tinv(0.975,804); *n-k-2 = 814-8-2 = 804;
run;
proc print data=work.tcrit;
run;
**crit value = 1.96292;

*leverage cut-off = 2(k+1)/n = 2(8+1)/804 = 0.022388;

*check for outliers and influential values;
proc print data=work.regdata2;
      where abs(jackknife) > 1.96292 or cooksD > 1 or leverage >0.022388;
run;

title 'jackknife';
proc print data=work.regdata2;
      where abs(jackknife) > 1.96292;
run;

title 'cooksD';
proc print data=work.regdata2;
      where cooksD > 1;
run;

title 'leverage';
proc print data=work.regdata2;
      where leverage >0.022388;
run;
```

5. 'thesis_index_sickle_aug19'

This file contains code for regression analysis of the relationship between index and sickle cell. This file also contains means analysis of the statistically significant interaction term between index and sickle cell and underweight.

```
*********************************;
* Kiersten Derby                 *;
* Thesis Regression Analysis  *;
* Feb 6, 2012                          *;
*********************************;

*-----------------------------------------------------------;
*Call in 'samplesize' data from thesis_datasteps.sas    ;
*-----------------------------------------------------------;

%include "H:\Thesis\thesis_datasteps_aug19.sas";

*-------------------------------------------------------------;
*Regression Analysis: check for interaction and confounding   ;
*-------------------------------------------------------------;
***********************;
* Index and Sickle    *;
***********************;

*Check for linearity for continuous variables age and RBP with TfR;
title;
ods listing;
symbol2 interpol=rl value=star color=blue;
proc gplot data=samplesize;
      plot log_index*rbp log_index*childagemonths_lab;
run;

*-----;
* AGE ;
*-----;
*Interaction: Test for interaction and confounding between sickleYN and age;
proc surveyreg data=samplesize;
      class sickleYN;
      cluster cluster;
      model log_index = sickleYN childagemonths_lab
sickleYN*childagemonths_lab / solution;
      format sickleYN sickleYN.;
run;

*confounding: method 2 (betas);
*SLR for crude beta;
proc surveyreg data=samplesize;
      class sickleYN;
      cluster cluster;
      model log_index = sickleYN / solution;
      format sickleYN sickleYN.;
run;

*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN;
      cluster cluster;
      model log_index = sickleYN childagemonths_lab / solution;
      format sickleYN sickleYN.;
```

```
run;


*-----;
* SEX ;
*-----;
*Interaction: Test for interaction and confounding between sickleYN and sex;
proc surveyreg data=samplesize;
      class sickleYN sex1;
      model log_index = sickleYN sex1 sickleYN*sex1 / solution;
      cluster cluster;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN sex1; *reference sex is 1 (female);
      cluster cluster;
      model log_index = sickleYN sex1 / solution;
      format sickleYN sickleYN.;
run;


*---------;
* Inflamm ;
*---------;
*Inflammation: Test for interaction and confounding between sickleYN and
inflammation;
proc surveyreg data=samplesize;
      model log_index = sickleYN inflammation1 sickleYN*inflammation1 /
solution;
      cluster cluster;
      format sickleYN sickleYN. inflammation1 yesno.;
run;

*confounding: method 2 (betas);

*MLR for adjusted beta;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_index = sickleYN inflammation1 / solution;
      format sickleYN sickleYN. inflammation1 yesno.;
run;

*----;
*RBP ;
*----;
*Interaction: Test for interaction and confounding between sickleYN and RBP;
proc surveyreg data=samplesize;
      class sickleYN;
      model log_index = sickleYN rbp sickleYN*rbp / solution;
      cluster cluster;
      format sickleYN sickleYN.;
run;
```

```
*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN;
      cluster cluster;
      model log_index = sickleYN rbp / solution;
      format sickleYN sickleYN.;
run;


*---------;
*stunting ;
*---------;
*Interaction: Test for interaction and confounding between sickleYN and
stunting;
proc surveyreg data=samplesize;
      class sickleYN stunted;
      model log_index = sickleYN stunted sickleYN*stunted / solution;
      cluster cluster;
      format sickleYN sickleYN. stunted stunted.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN stunted;
      cluster cluster;
      model log_index = sickleYN stunted / solution;
      format sickleYN sickleYN. stunted stunted.;
run;


*---------;
*Wasting  ;
*---------;
*Interaction: Test for interaction and confounding between sickleYN and
stunting;
proc surveyreg data=samplesize;
      class sickleYN wasted;
      model log_index = sickleYN wasted sickleYN*wasted / solution;
      cluster cluster;
      format sickleYN sickleYN. wasted wasted.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN wasted;
      cluster cluster;
      model log_index = sickleYN wasted / solution;
      format sickleYN sickleYN. wasted wasted.;
run;
```

```
*-------------;
*Underweight  ;
*-------------;
*Interaction: Test for interaction and confounding between sickleYN and
underweight;
proc surveyreg data=samplesize;
      class sickleYN underwt;
      model log_index = sickleYN underwt sickleYN*underwt / solution;
      cluster cluster;
      format sickleYN sickleYN. underwt underwt.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN underwt;
      cluster cluster;
      model log_index = sickleYN underwt / solution;
      format sickleYN sickleYN. underwt underwt.;
run;

*-------------;
* SES         ;
*-------------;
*SES: Test for interaction and confounding between sickleYN and SES where 1=
quintiles 1 or 2, 0 = quintiles 3-5;
*Interaction;
proc surveyreg data=samplesize;
      class sickleYN poor;
      model log_index = sickleYN poor sickleYN*poor / solution;
      cluster cluster;
      format sickleYN sickleYN. poor poor.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN poor;
      cluster cluster;
      model log_index = sickleYN poor / solution;
      format sickleYN sickleYN. poor poor.;
run;

*-------------;
* maternal edu;
*-------------;

*Maternal Education: Test for interaction and confounding between sickleYN
and education;
*Interaction;
proc surveyreg data=samplesize;
      model log_index = sickleYN education sickleYN*education / solution;
      cluster cluster;
      format sickleYN sickleYN. education yesno.;
```

```
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN education;
      cluster cluster;
      model log_index = sickleYN education / solution;
      format sickleYN sickleYN. education yesno.;
run;

*-------------;
* recent tea  ;
*-------------;
*Interaction: Test for interaction and confounding between sickleYN and tea;
proc surveyreg data=samplesize;
      class sickleYN tea24;
      model log_index = sickleYN tea24 sickleYN*tea24 / solution;
      cluster cluster;
      format sickleYN sickleYN. tea24 yesno.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN tea24;
      cluster cluster;
      model log_index = sickleYN tea24 / solution;
      format sickleYN sickleYN. tea24 yesno.;
run;

*-------------;
* Sprinkles;
*-------------;

*Interaction: Test for interaction and confounding between sickleYN and
sprinkles;
proc surveyreg data=samplesize;
      class sickleYN sprinkles24;
      model log_index = sickleYN sprinkles24 sickleYN*sprinkles24 / solution;
      cluster cluster;
      format sickleYN sickleYN. sprinkles24 yesno.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN sprinkles24;
      cluster cluster;
      model log_index = sickleYN sprinkles24 / solution;
      format sickleYN sickleYN. sprinkles24 yesno.;
run;

*-------------;
```

```
* Malaria     ;
*-------------;

*Interaction: Test for interaction and confounding between sickleYN and
malaria;
proc surveyreg data=samplesize;
      class sickleYN posmalaria;
      model log_index = sickleYN posmalaria sickleYN*posmalaria / solution;
      cluster cluster;
      format sickleYN sickleYN. posmalaria yesno.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN posmalaria;
      cluster cluster;
      model log_index = sickleYN posmalaria / solution;
      format sickleYN sickleYN. posmalaria yesno.;
run;

*-------------;
* Recent fever;
*-------------;

*Fever in 24 h: Test for interaction and confounding between sickleYN and
fever;
proc surveyreg data=samplesize;
      class sickleYN fever24;
      model log_index = sickleYN fever24 sickleYN*fever24 / solution;
      cluster cluster;
      format sickleYN sickleYN. fever24 yesno.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN fever24;
      cluster cluster;
      model log_index = sickleYN fever24 / solution;
      format sickleYN sickleYN. fever24 yesno.;
run;

*-----;
* Thal ;
*-----;
*Interaction: Test for interaction and confounding between sickleYN and
thalassemia;
proc surveyreg data=samplesize;
      class sickleYN;
      cluster cluster;
      model log_index = sickleYN thalYN sickleYN*thalYN / solution;
      format sickleYN sickleYN. thalYN yesno.;
run;
```

```
*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN;
      cluster cluster;
      model log_index = sickleYN thalYN / solution;
      format sickleYN sickleYN. thalYN yesno.;
run;

*-----;
*G6PD ;
*-----;
*Interaction: Test for interaction and confounding between sickleYN and g6pd;
proc surveyreg data=samplesize;
      class sickleYN;
      cluster cluster;
      model log_index = sickleYN g6pd_def sickleYN*g6pd_def / solution;
      format sickleYN sickleYN. g6pd_def yesno.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class sickleYN;
      cluster cluster;
      model log_index = sickleYN g6pd_def / solution;
      format sickleYN sickleYN. g6pd_def yesno.;
run;


*---------------------------;
* Exhaustive model          ;
*---------------------------;

*include all terms that were found to be significant for interaction or
confounding;
*interaction terms = underweight*sickle SES*sickle .
*sig in interaction = childagemonths_lab inflammation1 rbp wasted underwt
posmalaria
      fever24
*confounders = inflammation1 stunted poor education tea24 sprinkles24
posmalaria fever24
      thalYN g6pd_def;

proc surveyreg data=samplesize;
      cluster cluster;
      model log_index = sickleYN underwt underwt*sickleYN poor poor*sickleYN
childagemonths_lab inflammation1 rbp
            wasted posmalaria fever24 stunted education tea24 sprinkles24
thalYN g6pd_def/ solution;
      format sickleYN sickleYN. wasted wasted. poor poor. inflammation1
yesno. sprinkles24 yesno. posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.
```

```
tea24 yesno. thalYN yesno. g6pd_def yesno.;
run;

*----------------------------;
* Reduced model              ;
*----------------------------;

*take out sprinkles;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_index = sickleYN underwt underwt*sickleYN poor poor*sickleYN
childagemonths_lab inflammation1 rbp
            wasted posmalaria fever24 stunted education tea24 thalYN
g6pd_def/ solution;
      format sickleYN sickleYN. wasted wasted. poor poor. inflammation1
yesno. posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.
tea24 yesno. thalYN yesno. g6pd_def yesno.;
run;

*Take out education;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_index = sickleYN underwt underwt*sickleYN poor poor*sickleYN
childagemonths_lab inflammation1 rbp
            wasted posmalaria fever24 stunted tea24 thalYN g6pd_def/
solution;
      format sickleYN sickleYN. wasted wasted. poor poor. inflammation1
yesno. posmalaria yesno. stunted stunted.
            g6pd_def yesno. underwt underwt. fever24 yesno. tea24 yesno.
thalYN yesno. g6pd_def yesno.;
run;

*Take out thal;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_index = sickleYN underwt underwt*sickleYN poor poor*sickleYN
childagemonths_lab inflammation1 rbp
            wasted posmalaria fever24 stunted education tea24 sprinkles24
g6pd_def/ solution;
      format sickleYN sickleYN. wasted wasted. poor poor. inflammation1
yesno. sprinkles24 yesno. posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.
tea24 yesno. g6pd_def yesno.;
run;

*Take out RBP;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_index = sickleYN underwt underwt*sickleYN poor poor*sickleYN
childagemonths_lab inflammation1
            wasted posmalaria fever24 stunted education tea24 sprinkles24
thalYN g6pd_def/ solution;
      format sickleYN sickleYN. wasted wasted. poor poor. inflammation1
```

```
yesno. sprinkles24 yesno. posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.
tea24 yesno. thalYN yesno. g6pd_def yesno.;
run;

*Take out poor;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_index = sickleYN underwt underwt*sickleYN poor*sickleYN
childagemonths_lab inflammation1 rbp
            wasted posmalaria fever24 stunted education tea24 sprinkles24
thalYN g6pd_def/ solution;
      format sickleYN sickleYN. wasted wasted. inflammation1 yesno.
sprinkles24 yesno. posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.
tea24 yesno. thalYN yesno. g6pd_def yesno.;
run;

*Take out g6pd;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_index = sickleYN underwt underwt*sickleYN poor poor*sickleYN
childagemonths_lab inflammation1 rbp
            wasted posmalaria fever24 stunted education tea24 sprinkles24
thalYN / solution;
      format sickleYN sickleYN. wasted wasted. poor poor. inflammation1
yesno. sprinkles24 yesno. posmalaria yesno. stunted stunted.
            education yesno. underwt underwt. fever24 yesno. tea24 yesno.
thalYN yesno. g6pd_def yesno.;
run;

*Take out wasted;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_index = sickleYN underwt underwt*sickleYN poor poor*sickleYN
childagemonths_lab inflammation1 rbp
            posmalaria fever24 stunted education tea24 sprinkles24 thalYN
g6pd_def/ solution;
      format sickleYN sickleYN. poor poor. inflammation1 yesno. sprinkles24
yesno. posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.
tea24 yesno. thalYN yesno. g6pd_def yesno.;
run;

*take out tea;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_index = sickleYN underwt underwt*sickleYN poor poor*sickleYN
childagemonths_lab inflammation1 rbp
            wasted posmalaria fever24 stunted education sprinkles24 thalYN
g6pd_def/ solution;
      format sickleYN sickleYN. wasted wasted. poor poor. inflammation1
yesno. sprinkles24 yesno. posmalaria yesno. stunted stunted.
            education yesno. g6pd_def yesno. underwt underwt. fever24 yesno.
```

```
thalYN yesno. g6pd_def yesno.;
run;

*----------------------------;
* Final   model             ;
*----------------------------;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_index = sickleYN underwt underwt*sickleYN poor
childagemonths_lab inflammation1
             posmalaria fever24 stunted  thalYN g6pd_def/ solution;
      format sickleYN sickleYN. poor poor. inflammation1 yesno. sprinkles24
yesno. posmalaria yesno. stunted stunted.
            g6pd_def yesno. underwt underwt. fever24 yesno. thalYN yesno.
g6pd_def yesno.;
run;

*Reduced model with thal 3 levels - THIS IS FINAL MODEL;
proc surveyreg data=samplesize;
      class thal;
      cluster cluster;
      model log_index = sickleYN underwt underwt*sickleYN poor
childagemonths_lab inflammation1
             posmalaria fever24 stunted  thal g6pd_def/ solution;
      format sickleYN sickleYN. poor poor. inflammation1 yesno. sprinkles24
yesno. posmalaria yesno. stunted stunted.
            g6pd_def yesno. underwt underwt. fever24 yesno. thal thal.
g6pd_def yesno.;
run;

**add in other iron indicators to look at change in R-square (can only add ZP
because others used to define index**;
proc surveyreg data=samplesize;
      class thal;
      cluster cluster;
      model log_index = sickleYN underwt underwt*sickleYN poor
childagemonths_lab inflammation1
             posmalaria fever24 stunted  thal g6pd_def log_zp/ solution;
      format sickleYN sickleYN. poor poor. inflammation1 yesno. sprinkles24
yesno. posmalaria yesno. stunted stunted.
            g6pd_def yesno. underwt underwt. fever24 yesno. thal thal.
g6pd_def yesno.;
run;

*MEANS for INTERACTION TERM*;

*Sickle*underweight;
**Using proc surveymeans to account for cluster design**;
proc surveymeans data=samplesize;
   cluster cluster;
   domain sickleYN*underwt;
   var log_index;
   format sickleYN sickleYN. underwt underwt.;
run;
```

```
*************************;
*Check for collinearity   *;
*************************;

proc corr data=samplesize;
      var log_tfr sickleYN wasted underwt poor cf rbp posmalaria fever24
childagemonths_lab education sprinkles24
            thalYN g6pd_def;
      with sickleYN wasted underwt poor cf rbp posmalaria fever24
childagemonths_lab education sprinkles24 thalYN
            g6pd_def;
run;

proc reg data=work.samplesize plots(unpack);
      model log_tfr = thalYN cf posmalaria / partial vif;
run; quit;

*******************************************;
** Look at diagnostics of associative model **;
*******************************************;

ods graphics on;
ods exclude rfplot where= (_label_?'Intercept');

proc reg data=work.samplesize plots(unpack);
      model log_tfr = thalYN cf posmalaria / partial vif;
      output out=work.regdata2 R=resid P=yhat rstudent=jackknife cookd=cooksD
H=leverage;
run; quit;

proc print data=work.regdata2;
run;

*check normality on residuals;
proc univariate data=work.regdata2;
      var resid;
run;

**determine critical values;
data work.tcrit;
      tcrit = tinv(0.975,804); *n-k-2 = 809-3-2 = 804;
run;
proc print data=work.tcrit;
run;
**crit value = 1.96292;

*leverage cut-off = 2(k+1)/n = 2(3+1)/804 = 0.00995;
```

```
*check for outliers and influential values;
proc print data=work.regdata2;
      where abs(jackknife) > 1.96292 or cooksD > 1 or leverage >0.00995;
run;

title 'jackknife';
proc print data=work.regdata2;
      where abs(jackknife) > 1.96292;
run;

title 'cooksD';
proc print data=work.regdata2;
      where cooksD > 1;
run;

title 'leverage';
proc print data=work.regdata2;
      where leverage >0.00995;
run;

*35 observations with elevated jackknife. 0 obs for cooksD or leverage;
```

6. 'thesis_tfr_thal'

This file contains code for the regression analysis of the relationship between TfR and alpha-thalassemia

```
*********************************;
* Kiersten Derby                 *;
* Thesis Regression Analysis  *;
* Feb 6, 2012                              *;
*********************************;

*-------------------------------------------------------------;
*Call in 'samplesize' data from thesis_datasteps.sas    ;
*-------------------------------------------------------------;

%include "H:\Thesis\thesis_datasteps_aug19.sas";

*-------------------------------------------------------------------;
*Regression Analysis: check for interaction and confounding   ;
*-------------------------------------------------------------------;
***********************;
* Tfr AND Thal        *;
***********************;

*Check for linearity for continuous variables age and RBP with TfR;
title;
ods listing;
symbol2 interpol=rl value=star color=blue;
proc gplot data=samplesize;
```

```
      plot log_tfr*rbp log_tfr*childagemonths_lab;
run;

*-----;
* AGE ;
*-----;
*Interaction: Test for interaction and confounding between thal and age;
proc surveyreg data=samplesize;
      class thalYN;
      cluster cluster;
      model log_tfr = thalYN childagemonths_lab sickleYN*childagemonths_lab /
solution;
      format thalYN yesno.;
run;

*confounding: method 2 (betas);
*SLR for crude beta;
proc surveyreg data=samplesize;
      class thalYN;
      cluster cluster;
      model log_tfr = thalYN / solution;
      format thalYN yesno.;
run;

*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class thalYN;
      cluster cluster;
      model log_tfr = thalYN childagemonths_lab / solution;
      format thalYN yesno.;
run;


*-----;
* SEX ;
*-----;
*Interaction: Test for interaction and confounding between sickleYN and sex;
proc surveyreg data=samplesize;
      class thalYN sex1;
      model log_tfr = thalYN sex1 thalYN*sex1 / solution;
      cluster cluster;
      format thalYN yesno.;
run;
*no interaction between sickle cell and sex - now look for confounding
without interaction term in model;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class thalYN sex1; *reference sex is 1 (female);
      cluster cluster;
      model log_tfr = thalYN sex1 / solution;
      format thalYN yesno.;
run;
```

```
*---------;
* Inflamm ;
*---------;
*Inflammation: Test for interaction and confounding between sickleYN and
inflammation;
proc surveyreg data=samplesize;
      class thalYN cf;
      model log_tfr = thalYN cf thalYN*cf / solution;
      cluster cluster;
      format thalYN yesno. cf cf.;
run;

*Inflammation1 instead of cf*;
proc surveyreg data=samplesize;
      model log_tfr = thalYN inflammation1 thalYN*inflammation1 / solution;
      cluster cluster;
      format thalYN yesno. inflammation1 yesno.;
run;


*confounding: method 2 (betas);

*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class thalYN cf; *reference is cf=1 or reference;
      cluster cluster;
      model log_tfr = thalYN cf / solution;
      format thalYN yesno. cf cf.;
run;

*using 'inflammation1';
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = thalYN inflammation1 / solution;
      format thalYN yesno. inflammation1 yesno.;
run;

*----;
*RBP ;
*----;
*Interaction: Test for interaction and confounding between sickleYN and RBP;
proc surveyreg data=samplesize;
      class thalYN;
      model log_tfr = thalYN rbp thalYN*rbp / solution;
      cluster cluster;
      format thalYN yesno.;
run;
*no interaction between sickle cell and rbp - now look for confounding
without interaction term in model;

*confounding: method 2 (betas);
```

```
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class thalYN;
      cluster cluster;
      model log_tfr = thalYN rbp / solution;
      format thalYN yesno.;
run;


*---------;
*stunting ;
*---------;
*Interaction: Test for interaction and confounding between sickleYN and
stunting;
proc surveyreg data=samplesize;
      class thalYN stunted;
      model log_tfr = thalYN stunted thalYN*stunted / solution;
      cluster cluster;
      format thalYN yesno. stunted stunted.;
run;
*no interaction between sickle cell and rbp - now look for confounding
without interaction term in model;

*confounding: method 2 (betas);

*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class thalYN stunted;
      cluster cluster;
      model log_tfr = thalYN stunted / solution;
      format thalYN yesno. stunted stunted.;
run;


*---------;
*Wasting  ;
*---------;
*Interaction: Test for interaction and confounding between sickleYN and
stunting;
proc surveyreg data=samplesize;
      class thalYN wasted;
      model log_tfr = thalYN wasted thalYN*wasted / solution;
      cluster cluster;
      format thalYN yesno. wasted wasted.;
run;
*no interaction between sickle cell and wasting - now look for confounding
without interaction term in model;


*confounding: method 2 (betas);

*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class thalYN wasted;
```

```
      cluster cluster;
      model log_tfr = thalYN wasted / solution;
      format thalYN yesno. wasted wasted.;
run;

*-------------;
*Underweight  ;
*-------------;
*Interaction: Test for interaction and confounding between sickleYN and
stunting;
proc surveyreg data=samplesize;
      class thalYN underwt;
      model log_tfr = thalYN underwt thalYN*underwt / solution;
      cluster cluster;
      format thalYN yesno. underwt underwt.;
run;
*no interaction between sickle cell and wasting - now look for confounding
without interaction term in model;


*confounding: method 2 (betas);

*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class thalYN underwt;
      cluster cluster;
      model log_tfr = thalYN underwt / solution;
      format thalYN yesno. underwt underwt.;
run;

*-------------;
* SES         ;
*-------------;
*SES: Test for interaction and confounding between sickleYN and SES where 1=
quintiles 1 or 2, 0 = quintiles 3-5;
*Interaction;
proc surveyreg data=samplesize;
      class thalYN poor;
      model log_tfr = thalYN poor thalYN*poor / solution;
      cluster cluster;
      format thalYN yesno. poor poor.;
run;
*no interaction between sickle cell and wasting - now look for confounding
without interaction term in model;

*confounding: method 2 (betas);

*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class thalYN poor;
      cluster cluster;
      model log_tfr = thalYN poor / solution;
      format thalYN yesno. poor poor.;
run;
```

```
*-------------;
* maternal edu;
*-------------;

*Maternal Education: Test for interaction and confounding between sickleYN
and education;
*Interaction;
proc surveyreg data=samplesize;
      model log_tfr = thalYN education sickleYN*education / solution;
      cluster cluster;
      format thalYN yesno. education yesno.;
run;

*confounding: method 2 (betas);

*MLR for adjusted beta;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = thalYN education / solution;
      format thalYN yesno. education yesno.;
run;

*-------------;
* recent tea  ;
*-------------;

*Interaction: Test for interaction and confounding between sickleYN and tea;
proc surveyreg data=samplesize;
      class thalYN tea24;
      model log_tfr = thalYN tea24 thalYN*tea24 / solution;
      cluster cluster;
      format thalYN yesno. tea24 yesno.;
run;
*no interaction between sickle cell and tea - now look for confounding
without interaction term in model;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class thalYN tea24;
      cluster cluster;
      model log_tfr = thalYN tea24 / solution;
      format thalYN yesno. tea24 yesno.;
run;

*-------------;
* Sprinkles;
*-------------;

*Interaction: Test for interaction and confounding between sickleYN and
sprinkles;
proc surveyreg data=samplesize;
      class thalYN sprinkles24;
```

```
      model log_tfr = thalYN sprinkles24 thalYN*sprinkles24 / solution;
      cluster cluster;
      format thalYN yesno. sprinkles24 yesno.;
run;
*no interaction between sickle cell and sprinkles24 - now look for
confounding without interaction term in model;


*confounding: method 2 (betas);

*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class thalYN sprinkles24;
      cluster cluster;
      model log_tfr = thalYN sprinkles24 / solution;
      format thalYN yesno. sprinkles24 yesno.;
run;

*-------------;
* Malaria     ;
*-------------;

*Interaction: Test for interaction and confounding between sickleYN and
malaria;
proc surveyreg data=samplesize;
      class thalYN posmalaria;
      model log_tfr = thalYN posmalaria thalYN*posmalaria / solution;
      cluster cluster;
      format thalYN yesno. posmalaria yesno.;
run;
*no interaction between sickle cell and parasitemia - now look for
confounding without interaction term in model;

*confounding: method 2 (betas);

*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class thalYN posmalaria;
      cluster cluster;
      model log_tfr = thalYN posmalaria / solution;
      format thalYN yesno. posmalaria yesno.;
run;

*-------------;
* Recent fever;
*-------------;

*Fever in 24 h: Test for interaction and confounding between sickleYN and
fever;
proc surveyreg data=samplesize;
      class thalYN fever24;
      model log_tfr = thalYN fever24 thalYN*fever24 / solution;
      cluster cluster;
      format thalYN yesno. fever24 yesno.;
```

```
run;
*no interaction between sickle cell and fever - now look for confounding
without interaction term in model;

*confounding: method 2 (betas);

*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class thalYN fever24;
      cluster cluster;
      model log_tfr = thalYN fever24 / solution;
      format thalYN yesno. fever24 yesno.;
run;



*-----;
* sickle ;
*-----;

*Interaction: Test for interaction and confounding between sickleYN and
thalassemia;
proc surveyreg data=samplesize;
      class thalYN;
      cluster cluster;
      model log_tfr = thalYN sickleYN thalYN*sickleYN / solution;
      format thalYN yesno. sickleYN sickleYN.;
run;

*confounding: method 2 (betas);

*MLR for adjusted beta;
proc surveyreg data=samplesize;
      class thalYN;
      cluster cluster;
      model log_tfr = thalYN sickleYN / solution;
      format sickleYN sickleYN. thalYN yesno.;
run;

*-----;
*G6PD ;
*-----;

*Interaction: Test for interaction and confounding between sickleYN and g6pd;
proc surveyreg data=samplesize;
      class thalYN;
      cluster cluster;
      model log_tfr = thalYN g6pd_def thalYN*g6pd_def / solution;
      format thalYN yesno. g6pd_def yesno.;
run;

*confounding: method 2 (betas);

*MLR for adjusted beta;
```

```
proc surveyreg data=samplesize;
      class thalYN;
      cluster cluster;
      model log_tfr = thalYN g6pd_def / solution;
      format thalYN yesno. g6pd_def yesno.;
run;


*----------------------------;
* Exhaustive model          ;
*----------------------------;

*include all terms that were found to be significant for interaction or
confounding;
*no interaction terms.
* sig in interaction test: sex1 inflammation1 stunted underwt posmalaria
fever24
*confounders = recent tea consumption;

proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = thalYN sex1 inflammation1 stunted underwt posmalaria
fever24 tea24 / solution;
      format thalYN yesno. inflammation1 yesno.  posmalaria yesno. stunted
stunted.
            underwt underwt. fever24 yesno. tea24 yesno.;
run;

**Exhaustive model with SF and ZP**;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = thalYN sex1 inflammation1 stunted underwt posmalaria
fever24 tea24 log_sf log_zp / solution;
      format thalYN yesno. inflammation1 yesno.  posmalaria yesno. stunted
stunted.
            underwt underwt. fever24 yesno. tea24 yesno.;
run;

*----------------------------;
* Reduced model             ;
*----------------------------;

*take out underweight;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = thalYN sex1 inflammation1 stunted posmalaria fever24
tea24 / solution;
      format thalYN yesno. inflammation1 yesno.  posmalaria yesno. stunted
stunted.
            fever24 yesno. tea24 yesno.;
run;

*Take out fever;
proc surveyreg data=samplesize;
```

```
      cluster cluster;
      model log_tfr = thalYN sex1 inflammation1 stunted underwt posmalaria
tea24 / solution;
      format thalYN yesno. inflammation1 yesno.  posmalaria yesno. stunted
stunted.
            underwt underwt. tea24 yesno.;
run;

*Take out tea;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = thalYN sex1 inflammation1 stunted underwt posmalaria
fever24 / solution;
      format thalYN yesno. inflammation1 yesno.  posmalaria yesno. stunted
stunted.
            underwt underwt. fever24 yesno.;
run;

*Take out stunted;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = thalYN sex1 inflammation1 underwt posmalaria fever24
tea24 / solution;
      format thalYN yesno. inflammation1 yesno.  posmalaria yesno. underwt
underwt. fever24 yesno. tea24 yesno.;
run;

***reduced model with other iron markers;
*exhaustive;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = thalYN sex1 inflammation1 stunted underwt posmalaria
fever24 tea24 log_sf log_zp / solution;
      format thalYN yesno. inflammation1 yesno.  posmalaria yesno. stunted
stunted.
            underwt underwt. fever24 yesno. tea24 yesno.;
run;

*remove fever;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = thalYN sex1 inflammation1 stunted underwt posmalaria
tea24 log_sf log_zp / solution;
      format thalYN yesno. inflammation1 yesno.  posmalaria yesno. stunted
stunted.
            underwt underwt. fever24 yesno. tea24 yesno.;
run;

*remove underwt;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = thalYN sex1 inflammation1 stunted posmalaria fever24
tea24 log_sf log_zp / solution;
      format thalYN yesno. inflammation1 yesno.  posmalaria yesno. stunted
```

```
stunted.
            underwt underwt. fever24 yesno. tea24 yesno.;
run;

*remove tea;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = thalYN sex1 inflammation1 stunted underwt posmalaria
fever24 log_sf log_zp / solution;
      format thalYN yesno. inflammation1 yesno.  posmalaria yesno. stunted
stunted.
            underwt underwt. fever24 yesno. tea24 yesno.;
run;

*remove stunted;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = thalYN sex1 inflammation1 underwt posmalaria fever24
tea24 log_sf log_zp / solution;
      format thalYN yesno. inflammation1 yesno.  posmalaria yesno. stunted
stunted.
            underwt underwt. fever24 yesno. tea24 yesno.;
run;

*remove sex;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = thalYN inflammation1 stunted underwt posmalaria fever24
tea24 log_sf log_zp / solution;
      format thalYN yesno. inflammation1 yesno.  posmalaria yesno. stunted
stunted.
            underwt underwt. fever24 yesno. tea24 yesno.;
run;

*--------------------------;
* Final   model            ;
*--------------------------;

proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = thalYN sex1 inflammation1 stunted posmalaria fever24
tea24 / solution;
      format thalYN yesno. inflammation1 yesno.  posmalaria yesno. stunted
stunted.
            fever24 yesno. tea24 yesno.;
run;

****try with 'thal' instead of 'thalYN' and add HbSS and G6PD back into model
- THIS IS THE FINAL MODEL****;
proc surveyreg data=samplesize;
      class thal;
      cluster cluster;
      model log_tfr = thal g6pd_def sickleYN sex1 inflammation1 stunted
posmalaria fever24 tea24 / solution;
```

```
      format thal thal. g6pd_def yesno. sickleYN sickleYN. inflammation1
yesno.  posmalaria yesno. stunted stunted.
            fever24 yesno. tea24 yesno.;
run;

**add in other iron indicators (SF, ZP) to look at change in R-square;
proc surveyreg data=samplesize;
      class thal;
      cluster cluster;
      model log_tfr = thal g6pd_def sickleYN sex1 inflammation1 stunted
posmalaria fever24 tea24
            log_sf log_zp / solution;
      format thal thal. g6pd_def yesno. sickleYN sickleYN. inflammation1
yesno.  posmalaria yesno. stunted stunted.
            fever24 yesno. tea24 yesno.;
run;


***FINAL MODEL WITH OTHER IRON INDICATORS, with other blood do's***;
*thal 2 levels;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_tfr = thalYN g6pd_def sickleYN inflammation1 posmalaria
log_sf log_zp / solution;
      format thalYN thalYN. inflammation1 yesno.  posmalaria yesno. stunted
stunted.
            underwt underwt. fever24 yesno. tea24 yesno.;
run;
*thal 3 levels
proc surveyreg data=samplesize;
      class thal;
      cluster cluster;
      model log_tfr = thal g6pd_def sickleYN inflammation1 posmalaria log_sf
log_zp / solution;
      format thal thal. inflammation1 yesno.  posmalaria yesno. stunted
stunted.
            underwt underwt. fever24 yesno. tea24 yesno.;
run;

**************************;
*Check for collinearity   *;
**************************;

proc corr data=samplesize;
      var log_tfr sickleYN wasted underwt poor cf rbp posmalaria fever24
childagemonths_lab education sprinkles24
            thalYN g6pd_def;
      with sickleYN wasted underwt poor cf rbp posmalaria fever24
childagemonths_lab education sprinkles24 thalYN
            g6pd_def;
run;

proc reg data=work.samplesize plots(unpack);
      model log_tfr = thalYN cf posmalaria / partial vif;
```

```
run; quit;

*******************************************;
** Look at diagnostics of associative model **;
*******************************************;

ods graphics on;
ods exclude rfplot where= (_label_?'Intercept');

proc reg data=work.samplesize plots(unpack);
      model log_tfr = thalYN cf posmalaria / partial vif;
      output out=work.regdata2 R=resid P=yhat rstudent=jackknife cookd=cooksD
H=leverage;
run; quit;

proc print data=work.regdata2;
run;

*check normality on residuals;
proc univariate data=work.regdata2;
      var resid;
run;

**determine critical values;
data work.tcrit;
      tcrit = tinv(0.975,804); *n-k-2 = 809-3-2 = 804;
run;
proc print data=work.tcrit;
run;
**crit value = 1.96292;

*leverage cut-off = 2(k+1)/n = 2(3+1)/804 = 0.00995;

*check for outliers and influential values;
proc print data=work.regdata2;
      where abs(jackknife) > 1.96292 or cooksD > 1 or leverage >0.00995;
run;

title 'jackknife';
proc print data=work.regdata2;
      where abs(jackknife) > 1.96292;
run;

title 'cooksD';
proc print data=work.regdata2;
      where cooksD > 1;
run;

title 'leverage';
proc print data=work.regdata2;
      where leverage >0.00995;
run;

*35 observations with elevated jackknife. 0 obs for cooksD or leverage;
```

7. 'thesis_zp_g6pd'

This file contains code for the regression analysis of the relationship between ZP and G6PD deficiency among men.

```
*********************************;
* Kiersten Derby                 *;
* Thesis Regression Analysis     *;
* ZP and G6PD def in men         *;
* Sept 16, 2012                          *;
*********************************;

*-----------------------------------------------------------;
*Call in 'samplesize' data from thesis_datasteps.sas    ;
*-----------------------------------------------------------;

%include "H:\Thesis\thesis_datasteps_aug19.sas";

*************;
* ONLY MEN  *;
*************;

*only keep MEN for analysis*;
data men;
      set work.samplesize;
      if sex1=0;
run;
*n=429;

*check that we properly restricted to just men;
proc freq data=samplesize;
      tables sex1;
run;
proc freq data=men;
      tables sex1;
run;

*Set reference values;
proc format;
      value g6pd_def
            0 = "1 No G6PD"
            1 = "0 Yes G6pD"
      ; *No G6pD as reference;

      value sickleYN
            0 = "1 No Sickle"
            1 = "0 Yes Sickle"
      ; *No sickle as reference;

      value cf
            1 = "4 Reference"
```

```
            2 = "1 Incubation"
            3 = "2 Early Convalescence"
            4 = "3 Late Convalescence"
      ; *Reference as reference;

      value stunted
            0 = "1 Not stunted"
            1 = "0 Stunted"
      ; *No stunting as reference;

      value wasted
            0 = "1 Not wasted"
            1 = "0 Wasted"
      ; *No wasting as reference;

      value underwt
            0 = "1 Not underweight"
            1 = "0 Underweight"
      ; *Not underweight as reference;

      value poor
            0 = "1 Not poor"
            1 = "0 Poor"
      ; *Not poor as reference;

      value yesno
            0 = "2 No"
            1 = "1 Yes"
            . = "0 Missing"
      ; *NO as reference;

run;

proc freq data=men;
      tables g6pd_def;
      format g6pd_def g6pd_def.;
run;

*-------------------------------------------------------------;
*Regression Analysis: check for interaction and confounding   ;
*-------------------------------------------------------------;
**********************;
* ZP and G6PD         *;
**********************;

*Check for linearity for continuous variables age and RBP with ZP;
title;
ods listing;
symbol2 interpol=rl value=star color=blue;
proc gplot data=men;
      plot log_zp*rbp log_zp*childagemonths_lab;
run;

*-----;
```

```
* AGE ;
*-----;
*Interaction: Test for interaction and confounding between G6PD_def and age;
proc surveyreg data=men;
      class G6PD_def;
      cluster cluster;
      model log_zp = G6PD_def childagemonths_lab G6PD_def*childagemonths_lab
/ solution;
      format G6PD_def G6PD_def.;
run;

*confounding: method 2 (betas);
*SLR for crude beta;
proc surveyreg data=men;
      class G6PD_def;
      cluster cluster;
      model log_zp = G6PD_def / solution;
      format G6PD_def G6PD_def.;
run;

*MLR for adjusted beta;
proc surveyreg data=men;
      class G6PD_def;
      cluster cluster;
      model log_zp = G6PD_def childagemonths_lab / solution;
      format G6PD_def G6PD_def.;
run;


*---------;
* Inflamm ;
*---------;
*Inflammation: Test for interaction and confounding between G6PD_def and
inflammation;
proc surveyreg data=men;
      model log_zp = G6PD_def inflammation1 G6PD_def*inflammation1 /
solution;
      cluster cluster;
      format G6PD_def G6PD_def. inflammation1 yesno.;
run;

*confounding: method 2 (betas);

*MLR for adjusted beta;
proc surveyreg data=men;
      cluster cluster;
      model log_zp = G6PD_def inflammation1 / solution;
      format G6PD_def G6PD_def. inflammation1 yesno.;
run;

*----;
*RBP ;
*----;
*Interaction: Test for interaction and confounding between G6PD_def and RBP;
```

```
proc surveyreg data=men;
      class G6PD_def;
      model log_zp = G6PD_def rbp G6PD_def*rbp / solution;
      cluster cluster;
      format G6PD_def G6PD_def.;
run;
*no interaction between sickle cell and rbp - now look for confounding
without interaction term in model;

*confounding: method 2 (betas);

*MLR for adjusted beta;
proc surveyreg data=men;
      class G6PD_def;
      cluster cluster;
      model log_zp = G6PD_def rbp / solution;
      format G6PD_def G6PD_def.;
run;


*---------;
*stunting ;
*---------;
*Interaction: Test for interaction and confounding between G6PD_def and
stunting;
proc surveyreg data=men;
      class G6PD_def stunted;
      model log_zp = G6PD_def stunted G6PD_def*stunted / solution;
      cluster cluster;
      format G6PD_def G6PD_def. stunted stunted.;
run;
*no interaction between sickle cell and rbp - now look for confounding
without interaction term in model;

*confounding: method 2 (betas);

*MLR for adjusted beta;
proc surveyreg data=men;
      class G6PD_def stunted;
      cluster cluster;
      model log_zp = G6PD_def stunted / solution;
      format G6PD_def G6PD_def. stunted stunted.;
run;


*---------;
*Wasting  ;
*---------;
*Interaction: Test for interaction and confounding between G6PD_def and
stunting;
proc surveyreg data=men;
      model log_zp = G6PD_def wasted G6PD_def*wasted / solution;
      cluster cluster;
      format G6PD_def G6PD_def. wasted wasted.;
```

```
run;

*confounding: method 2 (betas);

*MLR for adjusted beta;
proc surveyreg data=men;
      cluster cluster;
      model log_zp = G6PD_def wasted / solution;
      format G6PD_def G6PD_def. wasted wasted.;
run;

*-------------;
*Underweight  ;
*-------------;
*Interaction: Test for interaction and confounding between G6PD_def and
underwt;
proc surveyreg data=men;
      model log_zp = G6PD_def underwt G6PD_def*underwt / solution;
      cluster cluster;
      format G6PD_def G6PD_def. underwt underwt.;
run;

*confounding: method 2 (betas);

*MLR for adjusted beta;
proc surveyreg data=men;
      cluster cluster;
      model log_zp = G6PD_def underwt / solution;
      format G6PD_def G6PD_def. underwt underwt.;
run;

*-------------;
* SES         ;
*-------------;
*SES: Test for interaction and confounding between G6PD_def and SES where 1=
quintiles 1 or 2, 0 = quintiles 3-5;
*Interaction;
proc surveyreg data=men;
      model log_zp = G6PD_def poor G6PD_def*poor / solution;
      cluster cluster;
      format G6PD_def G6PD_def. poor poor.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=men;
      cluster cluster;
      model log_zp = G6PD_def poor / solution;
      format G6PD_def G6PD_def. poor poor.;
run;

*-------------;
* maternal edu;
*-------------;
```

```
*Maternal Education: Test for interaction and confounding between G6PD_def
and education;
*Interaction;
proc surveyreg data=men;
      model log_zp = G6PD_def education G6PD_def*education / solution;
      cluster cluster;
      format G6PD_def G6PD_def. education yesno.;
run;

*confounding: method 2 (betas);

*MLR for adjusted beta;
proc surveyreg data=men;
      cluster cluster;
      model log_zp = G6PD_def education / solution;
      format G6PD_def G6PD_def. education yesno.;
run;

*-------------;
* recent tea  ;
*-------------;

*Interaction: Test for interaction and confounding between G6PD_def and tea;
proc surveyreg data=men;
      model log_zp = G6PD_def tea24 G6PD_def*tea24 / solution;
      cluster cluster;
      format G6PD_def G6PD_def. tea24 yesno.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=men;
      cluster cluster;
      model log_zp = G6PD_def tea24 / solution;
      format G6PD_def G6PD_def. tea24 yesno.;
run;

*-------------;
* Sprinkles;
*-------------;

*Interaction: Test for interaction and confounding between G6PD_def and
sprinkles;
proc surveyreg data=men;
      model log_zp = G6PD_def sprinkles24 G6PD_def*sprinkles24 / solution;
      cluster cluster;
      format G6PD_def G6PD_def. sprinkles24 yesno.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=men;
      cluster cluster;
```

```
      model log_zp = G6PD_def sprinkles24 / solution;
      format G6PD_def G6PD_def. sprinkles24 yesno.;
run;

*-------------;
* Malaria     ;
*-------------;

*Interaction: Test for interaction and confounding between G6PD_def and
malaria;
proc surveyreg data=men;
      model log_zp = G6PD_def posmalaria G6PD_def*posmalaria / solution;
      cluster cluster;
      format G6PD_def G6PD_def. posmalaria yesno.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=men;
      cluster cluster;
      model log_zp = G6PD_def posmalaria / solution;
      format G6PD_def G6PD_def. posmalaria yesno.;
run;

*-------------;
* Recent fever;
*-------------;

*Fever in 24 h: Test for interaction and confounding between G6PD_def and
fever;
proc surveyreg data=men;
      model log_zp = G6PD_def fever24 G6PD_def*fever24 / solution;
      cluster cluster;
      format G6PD_def G6PD_def. fever24 yesno.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=men;
      cluster cluster;
      model log_zp = G6PD_def fever24 / solution;
      format G6PD_def G6PD_def. fever24 yesno.;
run;

*-----;
* Thal ;
*-----;

*Interaction: Test for interaction and confounding between G6PD_def and
thalassemia;
proc surveyreg data=men;
      cluster cluster;
      model log_zp = G6PD_def thalYN G6PD_def*thalYN / solution;
      format G6PD_def G6PD_def. thalYN yesno.;
```

```
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=men;
      cluster cluster;
      model log_zp = G6PD_def thalYN / solution;
      format G6PD_def G6PD_def. thalYN yesno.;
run;

*-----;
*Sickle ;
*-----;

*Interaction: Test for interaction and confounding between G6PD_def and
sickle;
proc surveyreg data=men;
      cluster cluster;
      model log_zp = G6PD_def sickleYN G6PD_def*sickleYN / solution;
      format G6PD_def G6PD_def. sickleYN yesno.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=men;
      cluster cluster;
      model log_zp = G6PD_def sickleYN / solution;
      format G6PD_def G6PD_def. sickleYN yesno.;
run;


*---------------------------;
* Exhaustive model          ;
*---------------------------;

*include all terms that were found to be significant for interaction or
confounding;
*interaction terms = G6PD*fever24
*sig in interaction = childagemonths_lab inflammation1 rbp  posmalaria
*confounders = RBP underwt sprinkles24 posmalaria;

proc surveyreg data=men;
      cluster cluster;
      model log_zp = G6PD_def fever24 fever24*G6PD_def childagemonths_lab
inflammation1 rbp posmalaria
            underwt sprinkles24 / solution;
      format G6PD_def G6PD_def. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno.
            underwt underwt. fever24 yesno.;
run;

****Exhaustive model with SF and TFR****;
proc surveyreg data=men;
      cluster cluster;
```

```
      model log_zp = G6PD_def fever24 fever24*G6PD_def childagemonths_lab
inflammation1 rbp posmalaria
            underwt sprinkles24 log_sf log_tfr/ solution;
      format G6PD_def G6PD_def. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno.
            underwt underwt. fever24 yesno.;
run;

*---------------------------;
* Reduced model             ;
*---------------------------;
*Remove interaction term because p<0.01 in exhaustive model;
*take out insignificant terms sequentially;

*take out underweight;
proc surveyreg data=men;
      cluster cluster;
      model log_zp = G6PD_def fever24 childagemonths_lab inflammation1 rbp
posmalaria
             sprinkles24 / solution;
      format G6PD_def G6PD_def. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno.
            fever24 yesno.;
run;

*Take out sprinkles;
proc surveyreg data=men;
      cluster cluster;
      model log_zp = G6PD_def fever24 childagemonths_lab inflammation1 rbp
posmalaria
            underwt / solution;
      format G6PD_def G6PD_def. inflammation1 yesno. posmalaria yesno.
            underwt underwt. fever24 yesno.;
run;

*Take out fever;
proc surveyreg data=men;
      cluster cluster;
      model log_zp = G6PD_def childagemonths_lab inflammation1 rbp posmalaria
            underwt sprinkles24 / solution;
      format G6PD_def G6PD_def. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno.
            underwt underwt. ;
run;

*Take out inflammation;
proc surveyreg data=men;
      cluster cluster;
      model log_zp = G6PD_def fever24 childagemonths_lab rbp posmalaria
            underwt sprinkles24 / solution;
      format G6PD_def G6PD_def. sprinkles24 yesno. posmalaria yesno.
            underwt underwt. fever24 yesno.;
run;
```

```
****Reduced model with SF and TfR***;
*Full model;
proc surveyreg data=men;
      cluster cluster;
      model log_zp = G6PD_def fever24 fever24*G6PD_def childagemonths_lab
inflammation1 rbp posmalaria
            underwt sprinkles24 log_sf log_tfr/ solution;
      format G6PD_def G6PD_def. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno.
            underwt underwt. fever24 yesno.;
run;

*take out sprinkles;
proc surveyreg data=men;
      cluster cluster;
      model log_zp = G6PD_def fever24 fever24*G6PD_def childagemonths_lab
inflammation1 rbp posmalaria
            underwt log_sf log_tfr/ solution;
      format G6PD_def G6PD_def. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno.
            underwt underwt. fever24 yesno.;
run;

*take out inflammation;
proc surveyreg data=men;
      cluster cluster;
      model log_zp = G6PD_def fever24 fever24*G6PD_def childagemonths_lab rbp
posmalaria
            underwt sprinkles24 log_sf log_tfr/ solution;
      format G6PD_def G6PD_def. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno.
            underwt underwt. fever24 yesno.;
run;

*take out underwt;
proc surveyreg data=men;
      cluster cluster;
      model log_zp = G6PD_def fever24 fever24*G6PD_def childagemonths_lab
inflammation1 rbp posmalaria
             sprinkles24 log_sf log_tfr/ solution;
      format G6PD_def G6PD_def. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno.
            underwt underwt. fever24 yesno.;
run;

*take out fever;
proc surveyreg data=men;
      cluster cluster;
      model log_zp = G6PD_def fever24*G6PD_def childagemonths_lab
inflammation1 rbp posmalaria
            underwt sprinkles24 log_sf log_tfr/ solution;
      format G6PD_def G6PD_def. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno.
            underwt underwt. fever24 yesno.;
```

```
run;

*take out age;
proc surveyreg data=men;
      cluster cluster;
      model log_zp = G6PD_def fever24 fever24*G6PD_def inflammation1 rbp
posmalaria
            underwt sprinkles24 log_sf log_tfr/ solution;
      format G6PD_def G6PD_def. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno.
            underwt underwt. fever24 yesno.;
run;



*----------------------------;
* Final   model             ;
*----------------------------;

*same as exhaustive except took out interaction term*;
proc surveyreg data=men;
      cluster cluster;
      model log_zp = G6PD_def fever24 childagemonths_lab inflammation1 rbp
posmalaria
            underwt sprinkles24 / solution;
      format G6PD_def G6PD_def. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno.
            underwt underwt. fever24 yesno.;
run;

*FINAL MODEL: Reduced model with thal 3 levels;
proc surveyreg data=men;
      class thal;
      cluster cluster;
      model log_zp = G6PD_def thal sickleYN fever24 childagemonths_lab
inflammation1 rbp posmalaria
            underwt sprinkles24 / solution;
      format G6PD_def G6PD_def. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno.
            underwt underwt. fever24 yesno. thal thal. sickleYN yesno.;
run;

**Add in other iron indicators to look at change in R-square;
proc surveyreg data=men;
      class thal;
      cluster cluster;
      model log_zp = G6PD_def thal sickleYN fever24 childagemonths_lab
inflammation1 rbp posmalaria
            underwt sprinkles24 log_sf log_tfr log_index / solution;
      format G6PD_def G6PD_def. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno.
            underwt underwt. fever24 yesno. thal thal. sickleYN yesno.;
run;
```

```
***FINAL MODEL WITH SF AND TFR*** - added back thal and sickle;

proc surveyreg data=men;
      class thal;
      cluster cluster;
      model log_zp = G6PD_def thal sickleYN fever24 rbp posmalaria log_sf
log_tfr/ solution;
      format G6PD_def G6PD_def. thal thal. sickleYN sickleYN. inflammation1
yesno. sprinkles24 yesno. posmalaria yesno.
            underwt underwt. fever24 yesno.;
run;

*******************;
* ADJUST FOR MEN  *;
*******************;

*-----;
* SEX ;
*-----;
*Interaction: Test for interaction and confounding between G6PD_def and age;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = G6PD_def sex1 G6PD_def*sex1 / solution;
      format G6PD_def G6PD_def.;
run;

*confounding: method 2 (betas);
*SLR for crude beta;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = G6PD_def / solution;
      format G6PD_def G6PD_def.;
run;

*MLR for adjusted beta;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = G6PD_def sex1 / solution;
      format G6PD_def G6PD_def.;
run;


*-----;
* AGE ;
*-----;
*Interaction: Test for interaction and confounding between G6PD_def and age;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = G6PD_def childagemonths_lab G6PD_def*childagemonths_lab
/ solution;
      format G6PD_def G6PD_def.;
run;
```

```
*confounding: method 2 (betas);

*MLR for adjusted beta;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = G6PD_def childagemonths_lab / solution;
      format G6PD_def G6PD_def.;
run;

*---------;
* Inflamm ;
*---------;
*Inflammation: Test for interaction and confounding between G6PD_def and
inflammation;
proc surveyreg data=samplesize;
      model log_zp = G6PD_def inflammation1 G6PD_def*inflammation1 /
solution;
      cluster cluster;
      format G6PD_def G6PD_def. inflammation1 yesno.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = G6PD_def inflammation1 / solution;
      format G6PD_def G6PD_def. inflammation1 yesno.;
run;

*----;
*RBP ;
*----;
*Interaction: Test for interaction and confounding between G6PD_def and RBP;
proc surveyreg data=samplesize;
      model log_zp = G6PD_def rbp G6PD_def*rbp / solution;
      cluster cluster;
      format G6PD_def G6PD_def.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = G6PD_def rbp / solution;
      format G6PD_def G6PD_def.;
run;

*---------;
*stunting ;
*---------;
*Interaction: Test for interaction and confounding between G6PD_def and
stunting;
proc surveyreg data=samplesize;
      model log_zp = G6PD_def stunted G6PD_def*stunted / solution;
```

```
      cluster cluster;
      format G6PD_def G6PD_def. stunted stunted.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = G6PD_def stunted / solution;
      format G6PD_def G6PD_def. stunted stunted.;
run;


*---------;
*Wasting  ;
*---------;
*Interaction: Test for interaction and confounding between G6PD_def and
stunting;
proc surveyreg data=samplesize;
      model log_zp = G6PD_def wasted G6PD_def*wasted / solution;
      cluster cluster;
      format G6PD_def G6PD_def. wasted wasted.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = G6PD_def wasted / solution;
      format G6PD_def G6PD_def. wasted wasted.;
run;

*-------------;
*Underweight  ;
*-------------;
*Interaction: Test for interaction and confounding between G6PD_def and
underwt;
proc surveyreg data=samplesize;
      model log_zp = G6PD_def underwt G6PD_def*underwt / solution;
      cluster cluster;
      format G6PD_def G6PD_def. underwt underwt.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = G6PD_def underwt / solution;
      format G6PD_def G6PD_def. underwt underwt.;
run;

*-------------;
* SES         ;
*-------------;
```

```
*SES: Test for interaction and confounding between G6PD_def and SES where 1=
quintiles 1 or 2, 0 = quintiles 3-5;
*Interaction;
proc surveyreg data=samplesize;
      model log_zp = G6PD_def poor G6PD_def*poor / solution;
      cluster cluster;
      format G6PD_def G6PD_def. poor poor.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = G6PD_def poor / solution;
      format G6PD_def G6PD_def. poor poor.;
run;

*-------------;
* maternal edu;
*-------------;
*Maternal Education: Test for interaction and confounding between G6PD_def
and education;
*Interaction;
proc surveyreg data=samplesize;
      model log_zp = G6PD_def education G6PD_def*education / solution;
      cluster cluster;
      format G6PD_def G6PD_def. education yesno.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = G6PD_def education / solution;
      format G6PD_def G6PD_def. education yesno.;
run;

*-------------;
* recent tea  ;
*-------------;
*Interaction: Test for interaction and confounding between G6PD_def and tea;
proc surveyreg data=samplesize;
      model log_zp = G6PD_def tea24 G6PD_def*tea24 / solution;
      cluster cluster;
      format G6PD_def G6PD_def. tea24 yesno.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = G6PD_def tea24 / solution;
      format G6PD_def G6PD_def. tea24 yesno.;
run;
```

```
*-------------;
* Sprinkles;
*-------------;
*Interaction: Test for interaction and confounding between G6PD_def and
sprinkles;
proc surveyreg data=samplesize;
      model log_zp = G6PD_def sprinkles24 G6PD_def*sprinkles24 / solution;
      cluster cluster;
      format G6PD_def G6PD_def. sprinkles24 yesno.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = G6PD_def sprinkles24 / solution;
      format G6PD_def G6PD_def. sprinkles24 yesno.;
run;

*-------------;
* Malaria     ;
*-------------;
*Interaction: Test for interaction and confounding between G6PD_def and
malaria;
proc surveyreg data=samplesize;
      model log_zp = G6PD_def posmalaria G6PD_def*posmalaria / solution;
      cluster cluster;
      format G6PD_def G6PD_def. posmalaria yesno.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = G6PD_def posmalaria / solution;
      format G6PD_def G6PD_def. posmalaria yesno.;
run;

*-------------;
* Recent fever;
*-------------;
*Fever in 24 h: Test for interaction and confounding between G6PD_def and
fever;
proc surveyreg data=samplesize;
      model log_zp = G6PD_def fever24 G6PD_def*fever24 / solution;
      cluster cluster;
      format G6PD_def G6PD_def. fever24 yesno.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      cluster cluster;
```

```
      model log_zp = G6PD_def fever24 / solution;
      format G6PD_def G6PD_def. fever24 yesno.;
run;

*-----;
* Thal ;
*-----;

*Interaction: Test for interaction and confounding between G6PD_def and
thalassemia;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = G6PD_def thalYN G6PD_def*thalYN / solution;
      format G6PD_def G6PD_def. thalYN yesno.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = G6PD_def thalYN / solution;
      format G6PD_def G6PD_def. thalYN yesno.;
run;

*-----;
*Sickle ;
*-----;
*Interaction: Test for interaction and confounding between G6PD_def and
sickle;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = G6PD_def sickleYN G6PD_def*sickleYN / solution;
      format G6PD_def G6PD_def. sickleYN yesno.;
run;

*confounding: method 2 (betas);
*MLR for adjusted beta;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = G6PD_def sickleYN / solution;
      format G6PD_def G6PD_def. sickleYN yesno.;
run;


*---------------------------;
* Exhaustive model          ;
*---------------------------;

*include all terms that were found to be significant for interaction or
confounding;
*interaction terms = none
*sig in interaction = sex1 inflammation1 rbp  stunted underwt posmalaria
fever24
*confounders = sex1 inflammation1 stunted poor sprinkles24;
```

```
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = G6PD_def sex1 inflammation1 rbp  stunted underwt
posmalaria fever24 poor sprinkles24 / solution;
      format G6PD_def G6PD_def. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno.
            underwt underwt. fever24 yesno. stunted stunted. poor poor.;
run;

*--------------------------;
* Reduced model            ;
*--------------------------;
*take out insignificant terms sequentially;

*take out Sprinkles;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = G6PD_def sex1 inflammation1 rbp  stunted underwt
posmalaria fever24 poor / solution;
      format G6PD_def G6PD_def. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno.
            underwt underwt. fever24 yesno. stunted stunted. poor poor.;
run;

*Take out underwt;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = G6PD_def sex1 inflammation1 rbp  stunted posmalaria
fever24 poor sprinkles24 / solution;
      format G6PD_def G6PD_def. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno.
            underwt underwt. fever24 yesno. stunted stunted. poor poor.;
run;

*Take out stunted;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = G6PD_def sex1 inflammation1 rbp underwt posmalaria
fever24 poor sprinkles24 / solution;
      format G6PD_def G6PD_def. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno.
            underwt underwt. fever24 yesno. stunted stunted. poor poor.;
run;

*Take out fever;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = G6PD_def sex1 inflammation1 rbp  stunted underwt
posmalaria poor sprinkles24 / solution;
      format G6PD_def G6PD_def. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno.
            underwt underwt. fever24 yesno. stunted stunted. poor poor.;
run;
```

```
*Take out rbp;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = G6PD_def sex1 inflammation1  stunted underwt posmalaria
fever24 poor sprinkles24 / solution;
      format G6PD_def G6PD_def. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno.
            underwt underwt. fever24 yesno. stunted stunted. poor poor.;
run;

*take out ses;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = G6PD_def sex1 inflammation1 rbp  stunted underwt
posmalaria fever24 sprinkles24 / solution;
      format G6PD_def G6PD_def. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno.
            underwt underwt. fever24 yesno. stunted stunted. poor poor.;
run;


*---------------------------;
* Final   model             ;
*---------------------------;
proc surveyreg data=samplesize;
      cluster cluster;
      model log_zp = G6PD_def sex1 inflammation1 posmalaria / solution;
      format G6PD_def G6PD_def. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno.
            underwt underwt. fever24 yesno. stunted stunted. poor poor.;
run;

*FINAL MODEL: Reduced model with thal 3 levels and add sickle cell;
proc surveyreg data=samplesize;
      class thal;
      cluster cluster;
      model log_zp = G6PD_def sex1 inflammation1 posmalaria thal sickleYN /
solution;
      format G6PD_def G6PD_def. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno.
            underwt underwt. fever24 yesno. thal thal. sickleYN yesno.;
run;

**Add in other iron indicators to look at change in R-square;
proc surveyreg data=samplesize;
      class thal;
      cluster cluster;
      model log_zp = G6PD_def sex1 inflammation1 posmalaria thal sickleYN
log_sf log_tfr log_index / solution;
      format G6PD_def G6PD_def. inflammation1 yesno. sprinkles24 yesno.
posmalaria yesno.
            underwt underwt. fever24 yesno. thal thal. sickleYN yesno.;
run;
```