1. Introduction

This report adopts two approaches, linear and logistic regression, in estimating the socioeconomic determinants of a child's nutritional status within the country of Tanzania. Using data extracted from the 1992 Tanzania Demographic and Health Survey, containing data on a variety of socio-economic and health determinant measures of 4083 children. The dependent variable of this study, child nutritional status, will be represented by the weight for age z-score, coded as, and referred to throughout the essay as 'waz'. 'waz' is a measure of the child's weight at the time of the survey, standardised using the World Health Organisation (WHO) standard population.

This report has two aims, **1)** Use linear regression to identify which variables are important for explaining variation in the 'waz' variable. **2)** Use binary logistic regression to identify the variables important for highlighting severe malnourishment, classified as children with 'waz' values less than -2.

2. Outline of Methods

2.1 Linear Regression

The methods used within the linear regression analysis can be summarised as follows:

1. Dependent and independent variables explored using scatter and box plots to identify any patterns and trends.
2. Variable selection was carried out using the forward selection method. This was done manually using SPSS. For categorical variables with categories > 2, *k*-1 dummy variables were calculated (k = number of categories). Interpretation of these dummy variables become *the observed difference between the dummy and the reference category*. For all such categorical variables the first category was chosen as the reference category.
3. The manual forward selection method involves starting with the null model (no explanatory variables), adding each variable separately to find the 'best' variable determined by the following criteria: 1) The variable that causes the largest R-square increase, 2) And being significant at the 5% ($p < 0.05$) level. This identified variable is then added to the model and the process is repeated until all significant variables are added.
4. Using the variables identified by the forward selection an interaction term is hypothesised and fitted.
5. Model selection is carried out between the forward selection and interaction models to identify if the inclusion of the interaction term significantly increases the models fit using ANOVA test.
6. Model diagnostics are then performed on the selected final model to identify potential assumption violations of linear regression.

2.2 Logistic Regression

The methods used within the logistic regression can be summarised as follows:

1. Dependent and independent variables explored using bar plots and line graphs to identify any patterns and trends.
2. Variable selection was carried using automated Forward: LR, this uses a variable selection method based on the likelihood ratio. The inclusion of a variable in the model is based on the significance of the change in the log-likelihood.
3. From this forward selection model an interaction term was hypothesised and fitted.
4. Model selection is carried out between the forward selection and interaction models to determine whether the inclusion of the interaction term significantly improves model fit.

*All variables, the coding, and reference categories used within the study can be found in **Table 10** of the **Appendix**.*

3. Model Fitting

3.1 Linear Regression

3.1.1 Exploratory Plots

From the initial exploratory plots, it was identified the variable 'child's age (months)' may not have a linear relationship with the outcome variable. This can be seen in **Figure 1** within the **Appendix**. Similar trends have been identified by Alderman *et al* (2006), who identified a strong non-linear impact of child age on child nutrition status. It was thus hypothesised a quadratic term may be required for this variable. This was further investigated using a simple linear regression. Analysis of the residuals with a gaussian loess line fitted for better interpretation (Jacoby 2000), found within **Figure 2** of the **Appendix**, does suggest a non-linear pattern, with positive residuals associated with low and high values of child's age, whereas negative residuals closer to the centre. However, as this pattern of departure from zero was small, it was decided the quadratic term was not essential, keeping the simpler model.

3.1.2 Forward Selection

The forward selection method identified eleven predictor variables that caused a significant increase in R-squared. **Table 9** of the **Appendix** displays the variables entered at each step of the selection. These identified eleven predictors of 'waz' are; child's age, household wealth, baby size at birth, place of birth, region, fever two weeks prior to the survey, toilet facilities, diarrhoea two weeks prior to the survey, mother's occupation, mother's education, and female (gender).

3.1.3 Forward Selection Model Interpretation

The outputs of this forward selection model can be found below, in **Table 1**. These outputs suggest an estimated 12% of the variation in the dependent variable 'waz', is explained by the selected explanatory variables. Additionally, the significance of the reported F value suggests this model provides a better fit to the data compared to the null model.

The fitted regression model for the forward selection:

$$\hat{y} = -0.756 - 0.033\,cage + 0.210\,hhwealth1 + 0.294\,hhwealth2 - 0.253 fever - 0.461 babysz + 0.156 region2 + 0.136\,region3 - 0.148\,region4 + 0.035 region5 - 0.215\,region6 + 0.038 place1 + 0.259 place2 - 0.175\,toilet1 + 0.304\,toilet2 - 0.201\,diarr - 0.106 m\_occ1 + 0.262\,m\_occ2 + 0.122 meduc1 + 0.096 meduc2 + 0.078 female$$

Table 1. Output and Summary for the Forward Selection Model

Variable	Unstandardized coefficient	Significance
Constant	-0.756	0.001
Childs age (months)	-0.033	0.000
Hhwealth1	0.210	0.000
Hhewalth2	0.294	0.000
fever	-0.253	0.000
Size of baby at birth	-0.461	0.000
Region2	0.156	0.057
Region3	0.136	0.014
Region4	-0.148	0.062
Region5	0.035	0.611
Region6	-0.215	0.005
Place1	0.038	0.859
Palce2	0.259	0.227
Toilet1	-0.175	0.001
Toilet2	0.304	0.103
Diarrhoea	-0.201	0.000
M_occ1	-0.106	0.020
M_occ2	0.262	0.063
Meduc1	0.122	0.005
Meduc2	0.096	0.437
Female child	0.078	0.046
R	R-Square	Adjusted R-square
0.352	0.124	0.120
Mean Square	F	Sig.
45.107	28.750	0.000

The results of this multiple linear regression suggest, controlling for all other variables in the model respectively, a one month increase in a child's age decreases the expected 'waz' by 0.035, this change is significant. This may indicate older children of Tanzania have a worse nutritional status. Children from 'medium' and 'high' household wealth have respectively 0.106 and 0.321 higher expected 'waz' compared to children from a 'low' household wealth. Both observed differences were significant. This may suggest children from lower household wealth suffer from worse nutrition. Unsurprisingly, the results suggest that children who had fever two weeks prior to the survey have an expected 'waz' 0.253 lower compared to children who did not have fever. This observed difference was significant. A child who was a relatively small baby at birth has a decreased expected 'waz' of 0.461 compared to children who were of a normal size at birth. This observed difference is significant.

In terms of the regions of Tanzania, children from the northern highlands, lake and southern highlands regions have respectively 0.156, 0.136 and 0.035 higher expected 'waz' compared to children from the coastal region. Whereas children from the central and south regions show an expected 'waz' lower of 0.148 and 0.215 respectively, compared to children from the coastal region. Only the observed differences between the south and coastal and between the lake and coastal regions are significant. The 'place of birth' coefficients suggests when children are born at home or at the hospital the expected 'waz' is higher by 0.038 and 0.259 respectively, compared to children born at 'other'. These observed differences were not significant; however, the inclusion of this variable is based on the significantly increased R-square when the variable is added to the model.

Children with toilet facilities recorded as 'pit' have a lower expected 'waz' of 0.175 compared to children with no toilet facilities. Whereas children with toilet facilities recorded as 'flush' have an expected 'waz' higher by 0.304 compared to children with no toilet facilities. This may indicate the importance of improved toilet facilities in improving nutrition. However, only the observed difference between children with 'pit' toilet facilities and children with no toilet facilities was significant. Similar to the fever coefficient, as expected children with diarrhoea two weeks prior to the survey have an expected 'waz' 0.201 lower than children who did not have diarrhoea. This observed difference was significant. In terms of the effect of the mothers occupation on the child's nutritional status, the results suggest when the mothers occupation is a 'manual' job, the child has an expected 'waz' 0.106 lower compared to when the child's mother has no job. Whereas when the mother's occupation is 'non-manual', the child has an expected 'waz' 0.262 higher compared to when the child's mother has no job. Only the observed difference between 'manual' and no job is significant.

Interestingly, the results suggest that an improvement in the mother's education can improve the child's nutritional status. When the mother's education is equal to primary level or secondary and above, the child has a higher expected 'waz' of 0.122 and 0.096 respectively, compared to children whose mother's education is equal to below primary level. However, only the observed difference between 'primary' and 'below primary' was significant. Finally, the last coefficient 'female' suggests female children have an expected 'waz' 0.078 above male children, which may suggest female children have a better nutritional status. This observed difference was significant.

3.1.4 Interaction Model

From these identified variables an interaction term was fitted to improve the model fit. It was hypothesised there may be an interaction between the variables 'fever two weeks before survey' and 'household wealth' as the effect of fever on nutritional status may be differ for the three levels of household wealth. Children from higher household wealth may receive better treatment and may already be of better health before the fever. This interaction term was fitted to the forward selection model. Only the coefficients of the main effect and interaction effects can be found below in **Table 2.** The full table for the model can be found in **Table 11** of the **Appendix.**

The inclusion of the interaction term has not changed the signs of the coefficients of the main effects, however, in the presence of the interaction term these main effects can no longer be interpreted independently. The interpretation of these main effects suggests that for children **without fever** two weeks prior to the survey, children from 'medium' and 'high' household wealth have a higher expected 'waz' of 0.249 and 0.352 respectively, compared to children from 'low' household wealth. The 'fever' coefficient suggests that for children **with fever** two weeks prior to the survey, children from a 'low' household wealth have an expected 'waz'

0.186 lower compared to children without fever. Interpretation of the interaction coefficients can be confusing and so a table of predicted 'waz' values for the different interactions between fever and household wealth is shown in **Table 3**. This table shows that as household wealth increases the predicted 'waz' also increases for both children with fever and children without. On average children **with fever** from a medium household wealth have an estimated 'waz' 0.144 (-0.824 - -0.968) higher compared to children from low household wealth. Children **with fever** from high household wealth on average have an estimated 'waz' 0.204 (-0.764 - -0.968) higher compared to children from low household wealth. These values suggest there may be an interaction present between fever and household wealth.

Table 2. Output and Summary for the Main Effect and Interaction Effect of the Interaction Model

Variable	Unstandardized coefficient	Significance
Constant	-0.782	0.001
Hhwealth1	0.249	0.000
Hhwealth2	0.352	0.000
Fever	-0.186	0.000
FeverHhwealth1	-0.105	0.278
FeverHhwealth2	-0.148	0.104
R	R-Square	Adjusted R-square
0.353	0.125	0.120
Mean Square	F	Sig.
41.187	26.255	0.000

Table 3. Predicted Values of 'waz' for the Interaction Between Fever and Household Wealth

	Fever	
Household Wealth	Yes	No
Low	-0.782 + -0.186 = **-0.968**	**-0.782**
Medium	-0.782 + -0.186 + 0.249 + -0.105 = **-0.824**	-0.782 + 0.249 = **-0.533**
High	-0.782 + -0.186 + 0.352 + -0.148 = **-0.764**	-0.782 + 0.352 = **-0.43**

3.1.5 Model Selection

The inclusion of the interaction term was tested through an ANOVA to identify whether the inclusion of the interaction significantly improves the fit of the forward selection model. This is possible as the two models are nested. The outputs of this test can be found below in **Table 4**. Although the inclusion of the interaction term may suggest there is indeed an interaction between the effect of fever and household wealth on child nutrition, the addition of this interaction did not cause a significant increase in R-squared ($p = 0.281$, $p > 0.281$). This, in combination with both interaction terms not being significantwithin the model suggest it is sensible to remove this interaction term and keep the simpler, parallel slopes model identified through the forward selection.

Table 4. Output of the ANOVA Between the Forward Selection Model and Interaction Model

Model	R Square	R Square Change	F Change	Sig. F Change
1	0.124	0.124	28.750	0.000
2	0.125	0.001	1.269	0.281

3.1.6 Model Diagnostics

Regression diagnostics of the final model selected, the forward selection model (table 3), was carried out to check for assumption violations of linear regression. These violation assumptions were checked by plotting the residuals of the models in two different ways. A Residuals vs Fitted plot was used to check the linear assumption and a Normal P-P plot used to check the residuals are normally distributed. Model diagnostic plots can be found within **figure 3** of the **Appendix**. The Normal P-P plot shows no clear deviations and that the assumption of normality has not been violated. This can be observed as the points fall on the line and no distinguishable pattern is present. The residuals vs fitted plot with a loess line fitted, may show a pattern to suggest the presence of non-linearity. However, this pattern is not distinguished and thus it is sensible to conclude the assumption of linearity has not been violated. Additionally, the plot suggests the assumption of constant variance has not been violated. The residual plot may also indicate the presence of outliers. It may be acceptable to remove these outliers, however, for this model these outliers were not removed. In summary the model diagnostics suggest the assumptions of linear regression have not been violated and the results are reliable.

3.2 Logistic Regression

3.2.1 Forward Selection: LR

The forward selection method identified the final model in 8 steps. However, the model identified after step 6 had a larger value for the hosmer and lemeshow test as shown in **table 12** of the **Appendix**. This indicates this model to be a better fit for the data. Therefore, the variables selected within step 6 of the automated selection will be the variables used in the forward:LR model. The variables selected as significant predictors of severe malnutrition within step 6 were; Child's age, size of baby at birth, fever two weeks prior to survey, place of birth, household wealth, and, region.

3.2.2 Forward Selection: LR Model Interpretation

The outputs of the forward selection:LR model can be found below in **Table 5**. Additionally, the results of the Omnibus test of model coefficients can be found in **Table 13** of the **Appendix.** The results of the Omnibus test suggest the variables within the model are significant explainers of severe malnutrition as $p < 0.05$. However, the results of the Hosmer and Lemeshow test suggest the model is not a good fit for the data as $p < 0.05$.

The fitted logistic regression model for the forward:LR selection:

$$\begin{aligned} logit(\pi) = {} & -1.976 + 0.227975\, cage + 0.812\, babysz + 0.088\, diarr + 0.275\, fever \\ & - 0.605\, place1 - 0.920\, place2 - 0.221\, meduc1 - 0.457\, meduc2 \\ & + 0.1971 toilet1 - 0.135\, toilet2 - 0.251\, hhwealth1 - 0.550\, hhwealth2 \\ & - 0.121\, region1 - 0.205\, region2 + 0.322 region3 - 0.130 region4 \\ & + 0.352 region5 \end{aligned}$$

Table 5. Forward Selection : LR Model Output

Variable	B	Significance	Exp(B)
Childs age (months)	0.034	0.000	1.035
Size of baby(1)	0.747	0.000	2.110
Fever(1)	0.394	0.000	1.483
place		0.000	
Place(1)	-0.470	0.184	0.625
Place(2)	-0.789	0.027	0.454
householdwealth		0.000	
Household wealth(1)	-0.267	0.001	0.766
Household wealth(2)	-0.607	0.000	0.545
region		0.000	
Region(1)	-0.133	0.374	0.876
Region(2)	-0.213	0.028	0.808
Region(3)	0.295	0.028	1.343
Region(4)	-0.073	0.566	0.930
Region(5)	0.384	0.003	1.468
Constant	-0.804	0.030	0.448
Hosmer and Lemeshow Test	-2 Log Likelihood	Cox & Snell R Square	Nagelkerke R Sqaure
0.017	4756.076	0.063	0.089

Interpretation: Odds Scale

The results of the logistic regression suggest, controlling for all other variables in the model respectively, a one month increase in child's age increases the odds of severe malnutrition by 1.035, suggesting older children are at a higher risk of severe malnutrition. This is significant. A child with a low weight at birth has a 2.110 increase in the odds of severe malnutrition compared to children of normal birth weight. This observed difference was significant. Children with fever two weeks prior to the survey have 1.483 increase in the odds of severe malnutrition. As all these variables increase the odds of severe malnutrition they may be regarded as risk factors of severe malnutrition. The results of the logistic regression suggest the place of birth decreases the risk of severe malnutrition. Children born at home and children born at the hospital both decrease the odds of severe malnutrition by 0.625 and 0.454 respectively, compared to children born at 'other'. However, these observed differences were not significant. The results may indicate the effect of increased household wealth leads to a decrease in risk of severe malnutrition. Children from medium and high household wealth have respectively 0.766 and 0.545 decrease in the odds of severe malnutrition compared to children from low household wealth. Both observed differences are significant. In terms of the effect of the region the child is from, the results suggest children from the northern highlands, lake, and southern highlands have respectively 0.876, 0.808, and 0.930 decrease in the odds of severe malnutrition compared to children from the coastal region. Whereas children from the central and south regions have respectively 1.343 and 1.468 increase in the odds of severe malnutrition compared to children from the coastal region. The observed differences between the region's lake, southern highlands, and south compared with coastal regions were significant.

Model Interpretation: Marginal effects

Marginal effects were calculated for each variable using the following formula: $\pi(1-\pi)\beta x$. Where βx represents the coefficients of the different variables of the model. Taking the mean of this calculation for each variable gives the average marginal effect of that variable. The mean marginal effect for each variable can be found in **table 14** of the **Appendix**.

These results suggest, on average, a one month increase in child's age increases the probability of being severely malnourished by 0.67%. Children with a low birth weight are shown to have an increase of 14.83% in the probability of being severely malnourished compared to children with a normal birth weight. Children with fever two weeks prior to the survey have 7.79% increase in the probability of severe malnourishment compared to children without fever. As all these variables significantly increase the probability of severe malnutrition, they may be key areas to target when attempting to tackle child malnutrition. Conversely, children born at home and children born at hospital have respectively 9.44% and 15.77% decrease in the probability of severe malnourishment compared to children born at 'other'. The results of the logistic regression suggest increased household wealth is associated with a decrease in the probability of severe malnourishment. A child from a medium, and a child from a high household wealth have respectively 4.98% and 15.31% decrease in the probability of being severely malnourished compared to children from low household wealth. In regard to the marginal effect of the region the child is from, the effect of children from the northern highlands, lake and southern highlands have respectively 2.66%, 4.20%, and 1.43% decrease in the probability of severe malnourishment compared to children from the coastal region. Whereas the effect of children from the central and south regions have respectively 5.85% and 7.56% increase in the probability of severe malnourishment compared to children from the coastal region.

3.2.3 Interaction Model

To allow for more comparisons to be made between the linear and logistic regression and to further assess the effect, the same interaction as the linear regression method will be fitted to the logistic regression model. The output of the coefficients for the main effects and interaction term between fever and household wealth when fitted to the forward:LR model can be found below in **Table 6**, the full table for the model can be found in **Table 15** of the **Appendix..**

Interpretation of the interaction is done in terms of the probability of success outcome, in this case the probability of severe malnutrition for each category of household wealth and the presence of fever or no fever. These probabilities were calculated while holding all over variables in the model constant at their mean value. For categorical variables, the mean is calculated as the observed proportions at each level. These calculated probabilities are displayed below in **Table 7**. The equations used for each probability can be found in **Figure 4** of the **Appendix**.

Interpretation of these predicted probabilities indicate a clear pattern, when the child has fever, the probability of severe malnutrition decreases as household wealth increases. It can also be observed this decreasing of probability is not continuous. The decrease from low household wealth is relatively small when the child has fever, around a 3% decrease. Whereas the decrease in probability between low and high household wealth is significant, around a 12% decrease in probability of severe malnourishment. Additionally, the same pattern can be observed when the child does not have fever, the probability of severe malnutrition decreases, highlighting the impact of household wealth on child nutritional status.

3.2.4 Model Selection

This interaction model was tested in comparison with the simpler model produced through the forward selection method. This interaction was tested in terms of improvement of fit of the model. The results of these tests can be found below in **table 8**. As can be seen from these tables this interaction does improve the fit of the model, despite the interaction terms themselves not being significant. This may present a valid argument to keep this more complex interaction model. However, as indicated by the Hosmer and Lemeshow Significance, the model is still not a good fit for the data ($p < 0.05$).

Table 6. Outputs of the Main Effects and Interaction Terms When the Interaction is Added to the Forward:LR Model

Variable	B	Significance	Exp(B)
Fever(1)	0.322	0.001	1.380
householdwealth		0.000	
Household wealth(1)	-0.343	0.002	0.709
Household wealth(2)	-0.646	0.000	0.524
Fever by Household Wealth		0.524	
Fever by Household Wealth (1)	0.190	0.261	1.209
Fever by Household wealth (2)	0.095	0.610	1.100
Constant	-0.781	0.036	0.458
Hosmer and Lemeshow Test	-2 Log Likelihood	Cox & Snell R Square	Nagelkerke R Sqaure
0.029	4754.785	0.063	0.089

Table 7. MCA Table of the Probability of Severe Malnutrition for the Interaction Fever by Household Wealth

	Fever	
Household Wealth	Yes	No
Low	0.38	0.31
Medium	0.35	0.24
High	0.26	0.19

Table 8. Table Presenting the Output of the Goodness of Fit test between the Forward:LR and Interaction Models

Model	Hosmer and Lemeshow Significance	Omnibus Significance	Cox & Snell R Square	Nagelkerke R Square
1	0.017	0.000	0.063	0.089
2 - Interaction	0.029	0.000	0.063	0.089

4. Summary

In both linear and logistic regression, the forward selection method entered the variable 'child's age (in months)' first, suggesting that that variable is important in both estimating variation and explaining severe malnutrition. Additionally, both models identified the following variables as significant in analysing 'waz'; *size of baby, fever, place of birth, household wealth and region*. These may highlight key variables to further investigate as those that caused a decrease in nutritional status may be viewed as indicators of potential poor nutrition.

In both the linear and logistic model, the variable 'size of baby at birth' was identified as causing the greatest decrease in the expected 'waz' and greatest increase in probability of severe malnutrition. This may highlight the importance of proactive measures needed to combat nutritional deficiencies within the mother.

Within the linear model the variable identified as causing the greatest increase in expected 'waz' was the 'toilet' variable, recording the toilet facilities of each child. This may suggest implementation of flushable toilet access could greatly increase nutritional status. Within the logistic model the variable 'place of birth', more specifically children born within hospitals was identified as causing the greatest decrease in probability of severe malnutrition, closely followed by high household wealth which may indicate the nutritional differences present amongst children from different socio-economic backgrounds.

Comparing the effect of including the interaction term between fever and household wealth, the linear model did not have a significant improvement suggesting this interaction is not useful in modelling the variation in waz. However, in the logistic model the addition of the interaction term did improved the fit of the model, suggesting this interaction may be important in highlighting severe malnutrition.

The logistic regression model was found to be a poor fit for the data, suggesting the variables used within this study are not key variables in highlighting severe malnutrition. This is further supported by the classification table of the interaction logistic model, **table 16** of the **Appendix** where only 10.9% of severe malnutrition cases were correctly predicted. Similarly, the linear model was only able to explain 12% of the variation of 'waz', implying 88% of the variation is not explained by the model. It may be reasonable to conclude in both models there is significant room for improvement suggesting there are some key variables missing from the models. Additionally, this could be due to the complexity of modelling a complex socio-economic phenomenon such as malnutrition.

References

Alderman, H., H. Hoogeveen, and M. Rossi. 2006. "Reducing Child Malnutrition in Tanzania. Combined Effects of Income Growth and Program Interventions." Economics and Human Biology, 4 (1): 1-23.

Jacoby, W.G., 2000. Loess:: a nonparametric, graphical tool for depicting relationships between variables. *Electoral Studies*, *19*(4), pp.577-613.

Burchi, F., 2010. Child nutrition in Mozambique in 2003: the role of mother's schooling and nutrition knowledge. *Economics & Human Biology*, *8*(3), pp.331-345.

Appendix

Figure 1. Scatter plot and line graph of ‘waz’ by ‘cage’ (child’s age)

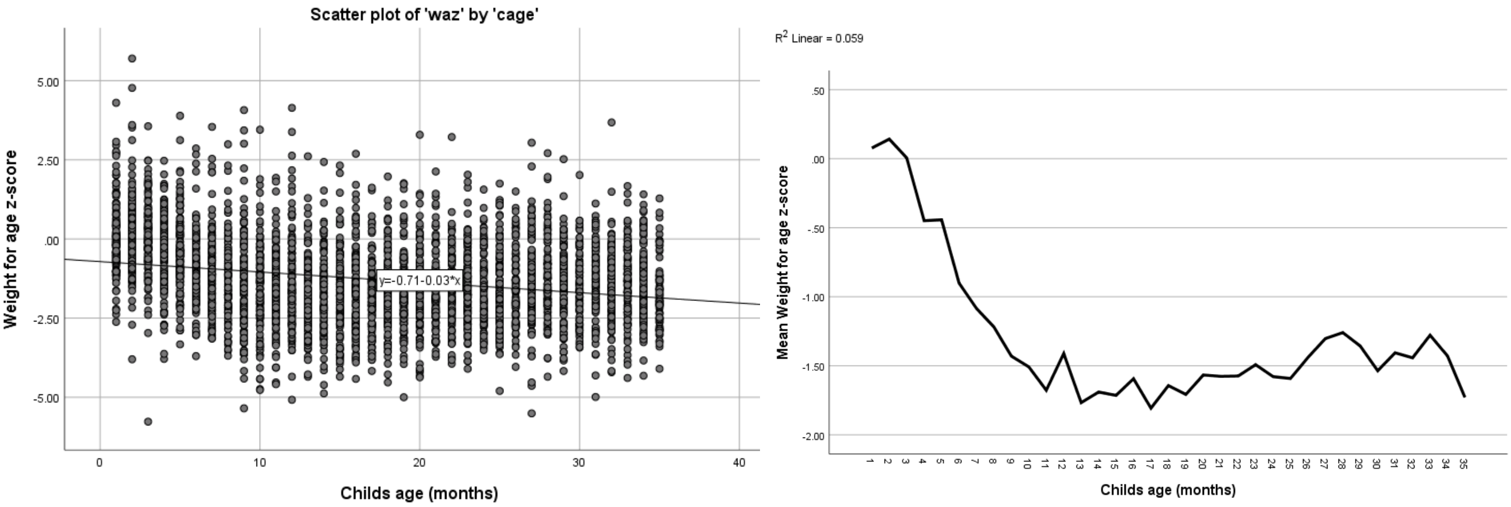

Figure 2. Scatter Plot of the Standardised Residuals by the Standardised Predicted Values for the Simple Linear Model of ‘waz’ and ‘cage’

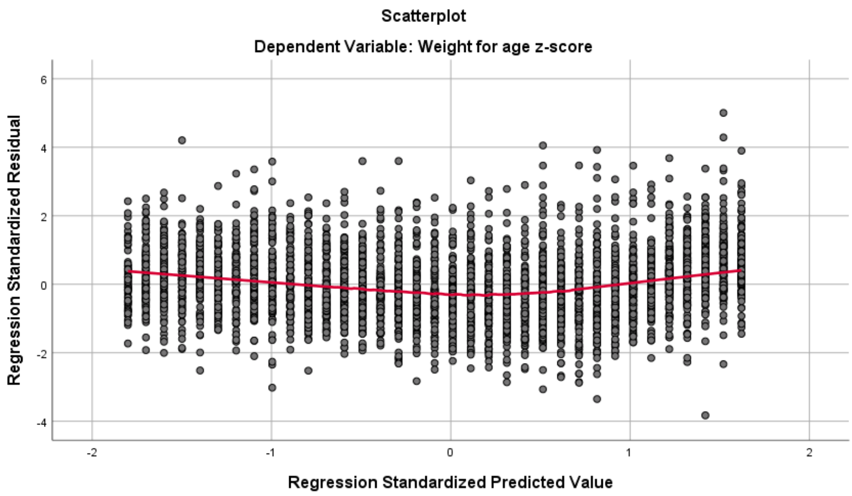

Table 9. Table Displaying the Variables Entered at Each Step of the Forward Selection Method of the Linear Model

Step	Variable Added	R squared increase
1	Cage	0.059
2	Hhwealth	0.018
3	Fever	0.011
4	babysz	0.008
5	Region	0.008
6	place	0.007
7	Toilet	0.004
8	Diarr	0.003
9	M_occ	0.003
10	Meduc	0.002
11	Female	0.001

Table 10. Complete List of Variables Used Within Analysis

Variable Name	Description	Categories	Reference Category
Dependent Variable			
Waz	Weight for age z-score	-	-
Independent variables			
Cage	Childs age (months)	-	-
Babysz	Size of baby at birth	0 = large/normal, 1 = small	'Large'
Preterm	Prematurity	0 = on time, 1 = premature	'On time'
Female	Female child	0 = male, 1 = female	'Male'
Diarr	Diarrhoea in 2 weeks prior to survey	0 = no, 1 = yes	'No'
Fever	Fever in 2 weeks prior to survey	0 = no, 1 = yes	'No'
Kidsu5	No. of children under 5 years in family	-	
Place	Place of Delivery	0 = other, 1= home, 2= hospital	'Other'
Pre_bi	Preceding birth interval	0 = 1^{st} birth, 1 = < 24 months, 2 = > 24 months	'1^{st} birth'
Meduc	Mother's education	0 = below primary, 1 = primary, 2 = secondary (or above)	'Below primary'
Peduc	Partner's education	0 = below primary, 1 = primary, 2 = secondary (or above)	'Below primary'
m_occ	Mother's occupation	0 = none, 1 = manual, 2 = non-manual	'None'
Toilet	Toilet facilities	0 = none, 1 = pit, 2 = flush	'None'
Hhwealth	Household Wealth	0 = low, 1 = medium, 2 = high	'Low'
Region	Region	1 = coastal, 2 = northern highlands, 3 = lake, 4 = central, 5 = southern highlands, 6 = south	'Coastal'
Religion	Respondent's Religion	0 = other, 1 = moslem, 2 = catholic, 3 = protestant, 4 = none	'Other'
Urbrur	Urban/Rural	0 = rural, 1 = urban, 2 = cities	'rural'

Figure 3. Normal P-P Plot and Residual vs Fitted Plot For The Linear Regression Model

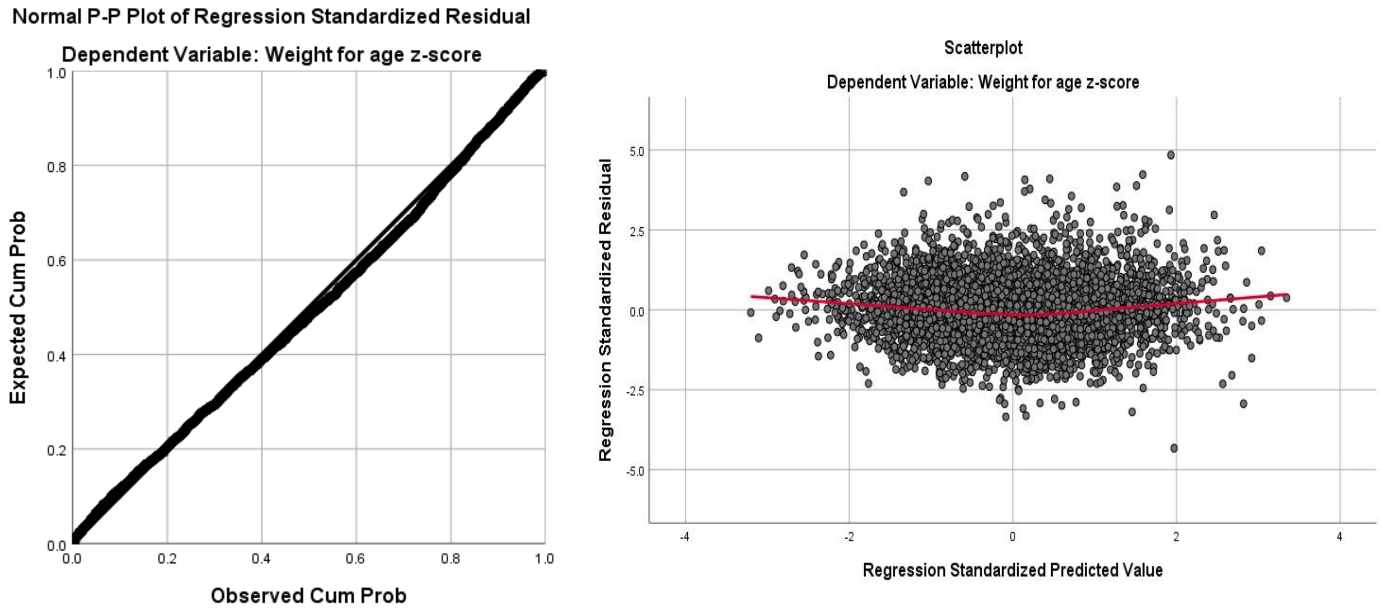

Table 11. Linear Regression Interaction Model Complete Output

Variable	Unstandardized coefficient	Significance
Constant	-0.782	0.001
Childs age (months)	-0.033	0.000
Hhwealth1	0.249	0.000
Hhewalth2	0.352	0.000
fever	-0.186	0.002
Size of baby at birth	-0.459	0.000
Region2	0.156	0.056
Region3	0.137	0.014
Region4	-0.145	0.069
Region5	0.039	0.580
Region6	-0.213	0.005
Place1	0.038	0.861
Palce2	0.259	0.228
Toilet1	-0.175	0.001
Toilet2	0.304	0.104
Diarrhoea	-0.201	0.000
M_occ1	-0.107	0.019
M_occ2	0.259	0.066
Meduc1	0.123	0.004
Meduc2	0.100	0.420
Female child	0.078	0.047
FeverHhwealth1	-0.105	0.278
FeverHhwealth2	-0.148	0.140
R	R-Square	Adjusted R-square
0.353	0.125	0.120
Mean Square	F	Sig.
45.197	26.255	0.000

Table 12. Logistic Regression Hosmer and Lemeshow Test for Each Step of the Forward:LR Selection

Step	Chi-square	df	Sig.
1	242.013	8	0.000
2	71.333	8	0.000
3	57.653	8	0.000
4	35.279	8	0.000
5	22.804	8	0.004
6	18.562	8	0.017
7	19.260	8	0.014
8	25.582	8	0.001

Table 13. Omnibus Test of Model Coefficients for the Forward: LR Model

Chi-square	df	sig
265.834	12	0.000

Table 14. Mean Marginal Effects for the Variables Within the Forward:LR Logistic Regression Model

	Variable											
	Cage	Babysz	Fever	Place1	Place2	Hhwealth1	Hhwealth2	Region2	Region3	Region4	Region5	Region6
Mean	0.0067	0.148	0.078	-0.093	-0.157	-0.053	-0.120	-0.026	-0.042	0.0585	-0.0145	0.0762

Table 15. Complete Table Output of Logistic Regression Interaction Model

Variable	B	Significance	Exp(B)
Childs age (months)	0.034	0.000	1.035
Size of baby(1)	0.745	0.000	2.107
Fever(1)	0.322	0.001	1.380
place		0.000	
Place(1)	-0.463	0.191	0.629
Place(2)	-0.783	0.028	0.457
householdwealth		0.000	
Household wealth(1)	-0.343	0.002	0.709
Household wealth(2)	-0.646	0.000	0.524
region		0.000	
Region(1)	-0.135	0.367	0.874
Region(2)	-0.214	0.027	0.807
Region(3)	0.294	0.029	1.342
Region(4)	-0.076	0.541	0.927
Region(5)	0.382	0.003	1.466
Fever by Household wealth		0.523	
Fever by Household wealth(1)	0.190	0.261	1.209
Fever by Household wealth(2)	0.095	0.610	1.100
Constant	-0.781	0.036	0.458
Hosmer and Lemeshow Test	-2 Log Likelihood	Cox & Snell R Square	Nagelkerke R Sqaure
0.029	4754.785	0.063	0.089

Figure 4. Equations Used To Calculate Probability for Interaction Term

Low household wealth and no fever

$$\pi = \frac{e^{-0.781+0.034*17.11+0.745*0.08-0.476*0.502-0.785*0.490-0.134*0.081-0.212*0.372+0.295*0.093-0.072*0.135+0.381*0.103}}{1+e^{-0.781+0.034*17.11+0.745*0.08-0.476*0.502-0.785*0.490-0.134*0.081-0.212*0.372+0.295*0.093-0.072*0.135+0.381*0.103}}$$

Low household wealth and fever

$$\pi = \frac{e^{-0.781+0.322+0.034*17.11+0.745*0.08-0.476*0.502-0.785*0.490-0.134*0.081-0.212*0.372+0.295*0.093-0.072*0.135+0.381*0.103}}{1+e^{-0.781+0.322+0.034*17.11+0.745*0.08-0.476*0.502-0.785*0.490-0.134*0.081-0.212*0.372+0.295*0.093-0.072*0.135+0.381*0.103}}$$

Medium household wealth and no fever

$$\pi = \frac{e^{-0.781-0.343+0.034*17.11+0.745*0.08-0.476*0.502-0.785*0.490-0.134*0.081-0.212*0.372+0.295*0.093-0.072*0.135+0.381*0.103}}{1+e^{-0.781-0.343+0.034*17.11+0.745*0.08-0.476*0.502-0.785*0.490-0.134*0.081-0.212*0.372+0.295*0.093-0.072*0.135+0.381*0.103}}$$

Medium household wealth and fever

$$\pi = \frac{e^{-0.781-0.343+0.322+0.190+0.034*17.11+0.745*0.08-0.476*0.502-0.785*0.490-0.134*0.081-0.212*0.372+0.295*0.093-0.072*0.135+0.381*0.103}}{1+e^{-0.781-0.343+0.322+0.190+0.034*17.11+0.745*0.08-0.476*0.502-0.785*0.490-0.134*0.081-0.212*0.372+0.295*0.093-0.072*0.135+0.381*0.103}}$$

High household wealth and no fever

$$\pi = \frac{e^{-0.781-0.646+0.034*17.11+0.745*0.08-0.476*0.502-0.785*0.490-0.134*0.081-0.212*0.372+0.295*0.093-0.072*0.135+0.381*0.103}}{1+e^{-0.781-0.646+0.034*17.11+0.745*0.08-0.476*0.502-0.785*0.490-0.134*0.081-0.212*0.372+0.295*0.093-0.072*0.135+0.381*0.103}}$$

High household wealth and fever

$$\pi = \frac{e^{-0.781-0.646+0.322+0.095+0.322+0.190+0.034*17.11+0.745*0.08-0.476*0.502-0.785*0.490-0.134*0.081-0.212*0.372+0.295*0.093-0.072*0.135+0.381*0.103}}{1+e^{-0.781-0.646+0.322+0.095+0.322+0.190+0.034*17.11+0.745*0.08-0.476*0.502-0.785*0.490-0.134*0.081-0.212*0.372+0.295*0.093-0.072*0.135+0.381*0.103}}$$

Table 16. Logistic Regression Interaction Model Classification Table

Severe Malnourishment	Percentage Correct
0.00 (no)	95.9
1.00 (yes)	11.2
Overall Percentage	70.0