World Journal of Public Health
2026; 11(1): 1-10
http://www.sciencepublishinggroup.com/j/wjph
doi: 10.11648/j.wjph.20261101.11

Hierarchical Modeling of Child Stunting Using Kenya Demographic Health Survey Data

Ombaka Ogolla[1,*], Richard Simwa[2], Cheruiyot Kipkoech[1]

[1]School of Mathematics and Actuarial Sciences, Technical University of Kenya, Nairobi, Kenya
[2]School of Science and Engineering, Department of Science, Daystar University, Nairobi, Kenya

Email address:

eombaka2000@gmail.com (Ombaka Ogolla), simwa@daystar.ac.ke (Richard Simwa),
Wycliffe.cheruiyot@tukenya.ac.ke (Cheruiyot Kipkoech)
*Corresponding author

To cite this article:

Ombaka Ogolla, Richard Simwa, Cheruiyot Kipkoech. (2026). Hierarchical Modeling of Child Stunting Using Kenya Demographic Health Survey Data. *World Journal of Public Health, 11*(1), 1-10. https://doi.org/10.11648/j.wjph.20261101.11

Received: 16 October 2025; **Accepted:** 5 November 2025; **Published:** 15 January 2026

Abstract: Child stunting reduction is the first of 6 goals in the Global Nutrition Targets for 2025 and a key indicator in the second Sustainable Development Goal of Zero Hunger. The prevalence of undernutrition is decreasing in many parts of the developing world, but challenges remain in many countries. For instance,the prevalence of stunting is 30.7% in Africa - higher than the global average of 22.0%. In Kenya, more than a quarter of children under the age of five, or two million children, have stunted growth. Stunting is the most frequent form of under-nutrition among young children. If not addressed, it has devastating long-term effects, including diminished mental and physical development.Child under-nutrition in Kenya has decreased in recent years. Levels of child stunting fell from 35.2% in 2009 to 26% in 2014 and wasting from 7% in 2009 to 4% in 2015. In Kenya, Coast Province has the highest stunting rate with (30.8%) and the lowest in Nairobi Province (17.2%). Despite this advancement, the world is still unlikely to achieve that goal in the global nutrition targets. Our study intends to investigate on crucial prognostic factors influencing child stunting in Coast, Kenya. The principal objective of this paper is to determine the effect of socioeconomic and demographic variables on child stunting in presence of dependencies in clusters and households. The study then uses variable selection technique, which is an artificial intelligence techniques to select covariates with the highest predictive power from the robust KDHS 2022 data. Additionally, a proportional hazards assumption test was carried out for the chosen covariates. Those covariates that satisfied the proportionality assumption were finally included in the frailty model to takes care of the presence of dependencies within the households. Data used were based on the Kenya Demographic and Health Survey (KDHS 2022), which were collected by use of questionnaires. Child stunting from the, KDHS 2022 data, was analyzed in an age period : stunting from the age of 12 months to the age of 60 months, referred to as “child stunting”.

Keywords: Stunting, Mixed Effects Models, Correlated Data, Function in R Statistical Software, Random Effects, Fixed Effects, Best Linear Unbiased Predictors (BLUP)

1. Introduction

Stunting is a process that can affect the development of a child from the early stages of conception until the third or fourth year of life, when the nutrition of the mother and the child are essential determinants of growth. Stunting is defined as the percentage of children whose ’height-for-age’ is below minus two standard deviations for moderate and minus three standard deviations for severe stunting from the median of the 2006 WHO Child Growth Standards[6].

Similarly, children are considered severely stunted if their length/height is below -3 SDs from the WHO child growth standards median for the same age and sex.

Nutritional stunting is caused by insufficient maternal nutrition, intrauterine undernutrition, lack of breastfeeding until 6 months of age, late introduction of complementary feeding, inadequate (quantity and quality) complementary feeding, and impaired absorption of nutrients owing to

infectious diseases[7, 8]. Stunting has long-term effects on individuals and societies, including poor cognition and educational performance, low adult wages, lost productivity and, when accompanied by excessive weight gain later in childhood, an increased risk of nutrition-related chronic diseases in adult life[9]. A high stunting rate generally signify unmet human health needs in education, medical care, nutrition and sanitation. The desire to understand the determinants of Child stunting poses a very important aspect of research. Today, it is an indicator in the second sustainable development goal of zero hunger. The Demographic and Health Surveys (DHS) program has been very instrumental in obtaining and disseminating authentic, national representative data on family planning, fertility, maternal and child health, among other health issues[20]. The most recent DHS survey conducted in Kenya was KDHS 2022.

This study aims at identifying the determinants of Child stunting in the Coastal region of Kenya. We chose a range of covariates obtained from three different publications based on Demographic health survey data in those three countries[1, 2, 11]. Those covariates were exceptional in determining child stunting. In the study, we then test for normality assumption using residual plots and analysis for purposes of diagnostics on the best fit from those specified covariates. We ultimately built a frailty model that takes care of the dependence within clusters and households.

A linear mixed effects model has two parts, that of fixed effects and random effects factors. A fixed effects factor is a factor whose levels are the only possible levels in the population being studied. This is opposed to a random effect factor whose levels in the study are just a sample of all the possible choices.

2. Methods

2.1. Data Description and Ethical Approval

In this paper, we use the Kenya Demographic Health Survey data *KDHS* 2022. It is the seventh Demographic and Health Survey (DHS) conducted in Kenya since 1989. KDHS is a national research undertaking conducted every five years with an intention of collecting a wide range of data with a strong interest on indicators of reproductive health, fertility, mortality, maternal and child health, nutrition and self-reported health habits among adults [24]. It is a household sample survey data with a national representation where households are selected at random from Kenya National Bureau of Statistics (KNBS) sampling frame. The survey procedures, instruments and sampling methods used in the KDHS 2022 acquired ethical recommendation from the Institutional Review Board of Opinion Research Corporation (ORC) Macro International Incorporated, a health, demographic, market research and consulting company situated in New Jersey, USA. We sought official registration on the DHS website and got permission to use the KDHS 2022 data. The data was downloaded in SPSS format and constituted 1,099 variables and 20,964 observations. Using package foreign, the data was imported to R software version 4.1.2 for analysis. KDHS data is a national survey data which is classified into 8 regions, constituting former provinces in Kenya. For this work, we analyzed data only for Coast region, being a region with the highest child stunting in Kenya. A set of dependant variables were chosen from literature given that they were profound in explaining child stunting. Survival time and status variables which are important considerations when analyzing survival data were calculated and included in the dataset.

2.2. Ethical Approval

The study did not need any approvals because it was secondary data.

3. Model Formulation

In our study, we discussed a series of models. First, was the *null_model*, without any covariate.

$$> null\_model < -glm(factor(Child\_is\_Stunted) \sim 1, family = binomial(link = "logit"), data = Child\_Stunt)$$

The second model is one with Cluster number as a random effect.

$$> fit1 < -glmer(factor(Child\_is\_Stunted) \sim (1 \mid Cluster\_number),$$
$$family = binomial(link = "logit"), data = Child\_Stunt)$$

A model with Household number as a random effect.

$$> fit2 < -glmer(factor(Child\_is\_Stunted) \sim (1 \mid Household\_number),$$
$$family = binomial(link = "logit"), data = Child\_Stunt)$$

Interaction model of Cluster and Household as a random effects.

$$> fit3 < -glmer(factor(Child\_is\_Stunted) \sim 1 + (1 \mid Cluster\_number/Household\_number),$$
$$family = binomial(link = "logit"), data = Child\_Stunt)$$

A model with fixed effects plus Cluster number as the random effect.

$$> fit4 < -glmer(Child\_is\_Stunted \sim as.factor(Place\_of\_residence) \\ +as.factor(Maternal\_Education\_level) + as.factor(Household\_wealth\_index) \\ +as.factor(Sex\_of\_child) + Age\_of\_child + Duration\_of\_breastfeeding \\ +as.factor(Frequency\_of\_listening\_to\_Radio) + as.factor(Size\_of\_child\_at\_birth) \\ +as.factor(Working\_status) + as.factor(Frequency\_of\_reading\_newspaper) \\ +as.factor(Frequency\_of\_watching\_TV) + as.factor(Source\_of\_drinking\_water) \\ +as.factor(Type\_of\_toilet\_facility) + (1 \mid Cluster\_number), \\ family = binomial(link = "logit"), data = Child\_Stunt$$

A model with fixed effects plus Household number as the random effect.

$$> fit5 < -glmer(Child\_is\_Stunted \sim as.factor(Place\_of\_residence) \\ +as.factor(Maternal\_Education\_level) + as.factor(Household\_wealth\_index) \\ +as.factor(Sex\_of\_child) + Age\_of\_child + Duration\_of\_breastfeeding \\ +as.factor(Frequency\_of\_listening\_to\_Radio) + as.factor(Size\_of\_child\_at\_birth) \\ +as.factor(Working\_status) + as.factor(Frequency\_of\_reading\_newspaper) \\ +as.factor(Frequency\_of\_watching\_TV) + as.factor(Source\_of\_drinking\_water) \\ +as.factor(Type\_of\_toilet\_facility) + (1 \mid Household\_number), \\ family = binomial(link = "logit"), data = Child\_Stunt$$

A model with fixed effects and interaction between random effects (Cluster with Household).

$$> fit6 < -glmer(Child\_is\_Stunted \sim as.factor(Place\_of\_residence) \\ +as.factor(Maternal\_Education\_level) + as.factor(Household\_wealth\_index) \\ +as.factor(Sex\_of\_child) + Age\_of\_child + Duration\_of\_breastfeeding \\ +as.factor(Frequency\_of\_listening\_to\_Radio) + as.factor(Size\_of\_child\_at\_birth) \\ +as.factor(Working\_status) + as.factor(Frequency\_of\_reading\_newspaper) \\ +as.factor(Frequency\_of\_watching\_TV) + as.factor(Source\_of\_drinking\_water) \\ +as.factor(Type\_of\_toilet\_facility) + (1 \mid Cluster\_number/Household\_number), \\ family = binomial(link = "logit"), data = Child\_Stunt$$

Since Household number is a random effect nested in Cluster number, we consider other model formulations bearing that in mind, before performing an ANOVA for comparison of the models. Consequently, choose the optimal one.

A nested model without any fixed effect covariate.

$$> fit7 < -glmer(Child\_is\_Stunted \sim (Cluster\_number \mid Household\_number), \\ family = binomial(link = "logit"), data = Child\_Stunt)$$

A nested model with all fixed and random effects.

$$> fit8 < -glmer(Child\_is\_Stunted \sim as.factor(Place\_of\_residence) \\ +as.factor(Maternal\_Education\_level) + as.factor(Household\_wealth\_index) \\ +as.factor(Sex\_of\_child) + Age\_of\_child + Duration\_of\_breastfeeding \\ +as.factor(Frequency\_of\_listening\_to\_Radio) + as.factor(Size\_of\_child\_at\_birth) \\ +as.factor(Working\_status) + as.factor(Frequency\_of\_reading\_newspaper) \\ +as.factor(Frequency\_of\_watching\_TV) + as.factor(Source\_of\_drinking\_water) \\ +as.factor(Type\_of\_toilet\_facility) + (Cluster\_number \mid Household\_number), \\ family = binomial(link = "logit"), data = Child\_Stunt$$

The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) information for model *fit4 and fit5* are the same both with AIC = 155.89, BIC = 261.28, Log Likelihood = -37.945. They are reasonably good models comparing the criterion values to that of the other model *fit6*. That model *fit6* had an AIC, BIC and p-value: 157.80, 265.83 and 0.7672 respectively. Unfortunately, we cannot compare *fit4 and fit5* directly with the other model fits since Bayesian Information Criterion for instance, dictates that we compare nested models[25].

It is recognised that smaller values for AIC, BIC define the best models.

$> anova(fit7, fit8, fit9)$

Table 1. *Comparison of the nested models.*

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
fit7	4	147.97	158.51	-69.985	139.97			
fit8	26	154.72	223.22	-51.358	102.72	37.253	22	0.02219 *
Signif.codes:	0	'***'	0.001	'**'	0.01	'*'	0.05	'.'

Comparatively, both models did not have a big disparity in their performance. Regardless of the fact that *fit7* which is unconditional model had a lower AIC and BIC values compared to model *fit8*. The model in *fit8* which was a saturated model had a great importance being statistically significant at $\alpha = 0.05$.

We recollect that an AIC is given by $AIC = -2\log L + 2k$, where $\log L$ is the maximum log-likelihood and k is the number of parameters in the model. Then BIC is given by $BIC = -2\log(L) + k\log(n)$ and for normally distributed errors, we have $BIC = \log(\sigma_\varepsilon^2) + \frac{k}{n}\log(n)$, where $\log L$ is the maximum log-likelihood, k is the number of parameters in the model, n is the number of observations and σ_ε^2 is the error variance. BIC is therefore an increasing function of error variance and the number of parameters[26].

By the "principle of parsimony" we choose the model

$$> fit9 < -glmer(Child\_is\_Stunted \sim as.factor(Sex\_of\_child) + as.factor(Education\_program) + as.factor(Number\_of\_children\_under\_5\_in\_HH) + as.factor(Number\_of\_births\_in\_last\_5\_years) + Maternal\_age\_at\_birth + as.factor(FP\_contraceptive\_use) + (Cluster\_number \mid Household\_number), family = binomial(link = "logit"), data = Child\_Stunt)$$

$$> fit10 < -lmer(Child\_is\_Stunted \sim as.factor(Sex\_of\_child) + as.factor(Education\_program) + as.factor(Number\_of\_children\_under\_5\_in\_HH) + as.factor(Number\_of\_births\_in\_last\_5\_years) + Maternal\_age\_at\_birth + as.factor(FP\_contraceptive\_use) + (Cluster\_number \mid Household\_number), data = Child\_Stunt, method = "REML")$$

$> anova(fit9, fit10)$

Table 2. *Comparison of the models.*

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
fit9	19	944.12	1057.76	-453.06	906.12			
fit10	20	-689.15	-569.53	364.58	-729.15	1635.3	1	<2.20E-16 ***
Signif. codes:	0.001	***						

In table 4 above, we made a comparison of the same covariates but under different criteria to understand their performance. Using the same rule of thumb that the best fit is the one with the lowest *AIC*, *fit10* is statistically significant with p-value <2.20E-16. Implying that it is a better model than a null model or even fit9.

3.1. Parameter Estimation

A linear mixed effects model is fitted of the form

$$Y_{ij} = \mu + \alpha_{ij} + \beta_{ij} + \varepsilon_{ij}$$

The matrix form for the mixed effects model is

$$Y = X\beta + Z\alpha + \varepsilon \quad (1)$$

where the vector β represents the fixed effect parameters, usually estimated by Generalised Least Squares (GLS) approach, the vector α represents the random effects and are estimated as Best Linear Unbiased Predictors (BLUPs). The mathematical theory of Mixed Model Analysis based on model (1) and well illustrated in literature such as Searle et. al., [27], requires that we estimate the parameters β and α. To estimate β, the fixed effects parameters, we use the method of GLS that maximizes the log-likelihood function with respect to β. We obtain Best Linear Unbiased Estimators (BLUEs) β via,

$$X'V^{-1}Y = X'V^{-1}X\beta \quad (2)$$

The GLS function (2) depends on the variance components via the matrix *V* and one has to obtain an estimate of matrix as a first step. This is done by REML (Restricted or Residual Maximum Likelihood) estimation. Then one needs the estimates of random effects. The estimates are referred to as "Predictors" to distinguish them from fixed effects for which the word "Estimates" has been used. The BLUPs of α are obtained from the equation,

$$\alpha = (Z'Z + \hat{\Gamma}^{-1})^{-1}Z'(Y - X\hat{\beta}) \quad (3)$$

In (3), the parameters to be estimated include Γ and β. The BLUE of β and the REML estimate of the variance components contained now in Γ, are substituted in the equation to finally obtain the BLUPs.

3.2. Restricted Maximum Likelihood Estimation

The variance components are two main parameters σ_α^2 and σ_ε^2 contained in the matrix *V*. The variance components can be estimated by a number of methods, including ANOVA, Maximum Likelihood, Bayesian Estimation and Method of Moments. Nevertheless, REML, developed by [28] is more attractive since it offers unbiased and non-negative estimates of variance components. Maximum likelihood estimates (MLE) of variance components may turn out to be negative [29]. Searle et al.[27], mentioned that such variance components can be set to zero.

4. Data Analysis

4.1. Descriptive Statistics

Table 3, children who were not stunted in Kilifi, Coast accounted for $123(35\%)$. This implied that more than half were stunted comprising of $224(65\%)$ and that should be a matter of concern and should be mitigated to further reduce it.

Table 3. *Child is stunted in Kilifi.*

Child stunted	Frequency	Percentage
No	123	0.35
Yes	224	0.65
Total	347	

In Table 4, the majority of the respondents resided in the rural areas of Kilifi in the Coast province with $272(70\%)$ while the rest are in the urban areas of Kilifi, which make up $113(30\%)$.

Table 4. *Place of residence.*

	Frequency	Percentage
Urban	113	0.30
Rural	272	0.70
Total	385	

Table 5, those with primary education comprised $228(59\%)$ followed by those who had Secondary education level $70(18\%)$ and the least were those with Higher education level $26(7\%)$.

Table 5. *Maternal Education Level.*

	Frequency	Percentage
No education	61	0.16
Primary	228	0.59
Secondary	70	0.18
Higher	26	0.07
Total	385	

Table 6, shows the distribution of children born in Coast province by gender. The numbers of males were slightly lower with $192(0.499\%)$ as compared to their female counterparts with $193(0.501\%)$.

Table 6. *Gender of children.*

	Frequency	Percentage
Male	192	0.499
Female	193	0.501
Total	385	

Table 7, shows the distribution of working status of mothers in Coast province. The numbers of those not working were higher with $241(0.63\%)$ as compared to their working female counterparts with $144(0.37\%)$.

Table 7. *Working Status.*

	Frequency	Percentage
No	241	0.63
Yes	144	0.37
Total	385	

In table 8, showed that children born with average size accounted for 81(70%). Those with very small size were 5(4.3%) while those with very large size at birth were 4(3%).

Table 8. *Size of child at birth.*

Size of child at birt	Frequency	Percentage
Very large	4	0.03
Larger than average	9	0.08
Average	81	0.70
Smaller than average	16	0.14
Very small	5	0.043
Total	115	

In Table 9, at least $200(52\%)$ had never watched a TV set, $50(13\%)$ of those interviewed had watched less than once a week and $135(35\%)$ watched at least once a week. Another category did not register even a single case was the 'almost every day'.

Table 9. *Frequency of watching TV.*

	Frequency	Percentage
Not at all	200	0.52
Less than once a week	50	0.13
At least once a week	135	0.35
Total	385	

4.2. Results of fitting Linear Mixed Model Using lmer

We consider the output for model from *fit10*. Our output below shows that REML was the tool that produced the variance estimates. It is the default tool under the *lmer* function, and if specified to be false, then *MLE* is used. The other information includes criterion for model choice, including *AIC*, *BIC* among others.

Linear mixed model fit by REML. t-tests use Satterthwaite's method [lmerModLmerTest]

$$> fit10 < -lmer(Child\_is\_Stunted \sim as.factor(Sex\_of\_child) + as.factor(Education\_program) + as.factor(Number\_of\_children\_under\_5\_in\_HH) + as.factor(Number\_of\_births\_in\_last\_5\_years) + Maternal\_age\_at\_birth + as.factor(FP\_contraceptive\_use) + (Cluster\_number \mid Household\_number), data = Child\_Stunt, method = "REML")$$

Table 10. *Models ANOVA output.*

AIC	BIC	logLik	$-2 * \log(L)$	df.resid
154.6	223.1	-51.3	102.6	77

Table 11. *Random effects.*

Groups	Name	Variance	Std.Dev.	Corr
Household_number	(Intercept)	2.243E-06	1.498E-03	
	Cluster_number	3.309E-10	1.819E-05	-1
Number of obs: 103	groups:	Household_number	62	

The output includes estimates of the variance components and standard deviations. The standard deviation of the variance component of Household_number is greater than that of the Cluster_number justifying its inclusion as a random effect. These variance estimates, which are elements of matrix *V* are then used to compute the fixed effect parameters via the formula (2).

Table 12. *Fixed effects.*

	Estimate	Std. Error	z value	Pr($>\|z\|$)
(Intercept)	1.36E+01	3.70E+01	0.367	0.7138
as.factor(Place_of_residence)2	1.67E+00	1.12E+00	1.496	0.1346
as.factor(Maternal_Education_level)1	9.03E-01	7.87E-01	1.147	0.2515
as.factor(Maternal_Education_level)2	6.33E-01	9.93E-01	0.637	0.5241
as.factor(Maternal_Education_level)3	1.90E+00	1.81E+00	1.051	0.2932
as.factor(Household_wealth_index)2	7.66E-01	9.22E-01	0.831	0.406

	Estimate	Std. Error	z value	Pr($>\|z\|$)
as.factor(Household_wealth_index)3	-6.30E-01	9.78E-01	-0.645	0.5192
as.factor(Household_wealth_index)4	8.76E-01	1.39E+00	0.63	0.5284
as.factor(Household_wealth_index)5	-1.80E+01	3.98E+03	-0.005	0.9964
as.factor(Sex_of_child)2	-4.21E-01	5.08E-01	-0.829	0.4072
Age_of_child	-8.61E-02	4.66E-01	-0.185	0.8533
Duration_of_breastfeeding	-1.70E-01	3.86E-01	-0.44	0.6599
as.factor(Frequency_of_listening_to_Radio)1	4.86E-01	7.65E-01	0.635	0.5255
as.factor(Frequency_of_listening_to_Radio)2	-4.91E-01	5.73E-01	-0.856	0.3923
as.factor(Size_of_child_at_birth)2	1.55E+00	1.68E+00	0.921	0.3571
as.factor(Size_of_child_at_birth)3	9.41E-01	1.46E+00	0.644	0.5198
as.factor(Size_of_child_at_birth)4	1.69E+00	1.61E+00	1.05	0.2938
as.factor(Size_of_child_at_birth)5	3.84E+01	6.71E+07	0	1
as.factor(Working_status)1	-4.98E-01	5.54E-01	-0.9	0.3681
as.factor(Frequency_of_reading_newspaper)1	-6.30E-01	1.02E+00	-0.615	0.5386
as.factor(Frequency_of_reading_newspaper)2	-2.80E+01	1.14E+06	0	1
as.factor(Frequency_of_watching_TV)1	1.99E-01	1.10E+00	0.18	0.8568
as.factor(Frequency_of_watching_TV)2	-1.37E+00	8.08E-01	-1.691	0.0909 .

The variable Sex of child had two levels namely, Male and Female. Male is the reference category. Consequently, female children had 34.36% less likely to be stunted as compared to those male children.

The children born in rural areas had a higher odd of being stunted by a factor of 5.31 as compared to their counterparts born in urban areas.

Those children born in household with mothers having higher education had higher odds of being stunted by a factor of 1.9 than those who did not have education at all.

The chances of child stunting was lower by a factor of 1.8 for children born in households with the richest wealth index as compared to the children born in households with the poorest wealth index.

As age of a child increases by one years, the chances of being stunted reduced by 8.2%.

As duration of breastfeeding of a child increases by one unit, the chances of being stunted reduced by 15%.

Interestingly, children who were born with very small size had a higher odd factor of 3.84 of being stunted as compared to those who were born with a very large size at birth which was the reference category.

Again, those parents who listened to radio at least once a week had 38% less likely chances of having their children stunted as compared to parents or families who did not listen to radio at all.

Families with working status had 39% less chances of having a stunted child as compared to those who were not working at all.

Families with the ability to access and read newspapers were at least once a week had a lower chances (by a factor of $6.9144001 \times 10^{-13}$) of having a stunted child as compared to those families who did not at all.

Families with the ability to watch a television set at least once a week had a lower chances by a factor of 74.5% of having a stunted child as compared to those families who did not at all.

4.3. Residual Analysis

Apart from the information on the exploratory data analysis, it is equally useful to consider the residual plots and analysis for purposes of diagnostics. Residual versus fitted value plots, normal QQ-plots, factor versus residual plots are some of the common tools in understanding the residual structure in the data. By using our $fit10$, we obtain the residual plot. Moreover, we employ a statistical inference method which is a more liberal test of normality done on the residuals of model $fit9$, which is the Shapiro Wilk test,

$$> shapiro.test(residuals(fit9))data : Child\_Stunt$$
$$residuals(fit9) \quad W = 0.95489,$$
$$p - value = 0.001457$$

From our result above, the normality assumption is violated by the results of Shapiro-Wilk test, i.e., the null hypothesis of normally distributed data is rejected as the p-value is significant at 5% level, $(p - value = 0.001457)$. The stem and leaf diagram gotten by $> stem(residuals(fit9))$ also confirms the non-normality assumption.

To establish the independence of the residuals and the homoskedasticity, we use a plot of the observed values versus residuals predicted values. Figure **??**, shows how normality assumption does not quite hold for the residuals. The points do not fall on the straight line violating that assumption. On the other plot, even though the points fall above and below the zero line. They are not randomly scattered with constant spread on that space for independence and homoskedasticity to hold. Secondly, the points are conjested at the zero line[26].

5. Discussion

The study attempts to understand the determinants of child stunting using survey data from Kenya Demographic Health

Survey. In this case, Kenya DHS survey 2022 dataset was used for the analysis. For this work, we analysed data for Coast province only being a region with the highest child stunting in Kenya.

A high rate of child stunting generally signify unmet human health needs in education, medical care, nutrition and sanitation. Therefore, aspiration to understand the determinants of Child stunting gives rise to a very important aspect of research. Many studies have employed regression techniques to explore the determinants of U5CM. The Cox PH regression was used by [30–32] among others.

In our findings using Linear Mixed Effects model we realised that the following covariates; Number of children under five years in household, Number of births in the last five years, Modern methods of family planning and contraceptive use were statistically significant in determining child stunting.

In comparison with other studies that were published, there was no big mismatch being that findings showed that child stunting was associated with variables related to: child characteristics at birth (such as age at birth), factors of reproduction of the mother (such as number of siblings born before), feeding characteristics and anthropometric measurements.

This was in line with other findings such as [30] who used Cox PH regression and established that region of residence, sex of the child, type of birth (multiple), birth interval (less than 24 months after the preceding birth), and mother's education were related with an increased risk of child stunting before their fifth birthday. [31] also established that factors related to mother characteristics and previous births such as sex of the child, sex of the head of the household and the number of births in the past one year was found to be significant. [32] explored the effect of mother's education, child's sex, rural/urban residence, household wealth index, regions ecological zones and development.

In our work, we had a large sample of 2926 observations and 1132 variables which was statistically sufficient. Nevertheless, there was a limitation of missing cases which we ignored and just used complete cases in our analysis.

Another limitation was that we did not use any scientific method on choosing the covariates we worked on. That, was just chosen from literature.

The study is of significance in that it will help the government, Non-governmental organisations to get information to guide them on how to intervene in order to ultimately achieve the global target on Child stunting.

Furthermore, students who are actively engaged in doing research in child stunting may not only fill the gap but also add on the already existing bank of knowledge.

6. Conclusion

Correlated data arise from numerous health fields especially where participants are in a cluster or household sharing the same prognostic factors. In this research, we have managed to present a framework for determination of child stunting using the 2022 KDHS data from Coast Province in Kenya.

Majorly, the framework involved checking which covariates were notable in explaining child stunting. Moreover, based on this research, child stunting is associated with variables related to factors of reproduction of the mother (such as number of children under five years in the household, number of births in the last five years and modern family planning methods like contraceptive use.

ORCID

0009-0004-8649-6752 (Ombaka Ogolla)
0000-0002-2136-1433 (Richard Simwa)

Abbreviations

ANOVA	Analysis of Variance
MLE	Maximum Likelihood Estimation
REML	Restricted Maximum Likelihood
lmer	Linear Mixed Effects Models

Acknowledgments

We acknowledge the anonymous reviewers whose remarks revamp the standard of this paper outstandingly.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Beal, T., Tumilowicz, A., Sutrisna, A., Izwardy, D., & Neufeld, L. M. (2018). A review of child stunting determinants in Indonesia. Maternal & child nutrition, 14(4), e12617.

[2] Hoffman, D., Cacciola, T., Barrios, P., & Simon, J. (2017). Temporal changes and determinants of childhood nutritional status in Kenya and Zambia. Journal of Health, Population and Nutrition, 36(1), 1-13.

[3] Global Nutrition report: https://globalnutritionreport.org/resources/nutrition-profiles/africa/ (2014)

[4] Unicef report: https://www.unicef.org/kenya/nutrition:2019

[5] Kenya National Bureau for Statistics report : KNBS. Kenya National Bureau for Statistics (KNBS) (2015) Demographic and health Survey 2014. Nairobi, Kenya: KNBS, Ministry of Health, NACC, KEMRI, NCPD, 2015.

[6] Soliman, A., De Sanctis, V., Alaaraj, N., Ahmed, S., Alyafei, F., Hamed, N., & Soliman, N. (2021). Early and long-term consequences of nutritional stunting: from childhood to adulthood. Acta Bio Medica: Atenei Parmensis, 92(1).

[7] Kramer, C. V., & Allen, S. (2015). Malnutrition in developing countries. Paediatrics and child health, 25(9), 422-427.

[8] Prendergast, A. J., & Humphrey, J. H. (2014). The stunting syndrome in developing countries. Paediatrics and international child health, 34(4), 250-265.

[9] Dewey, K. G., & Begum, K. (2011). Long-term consequences of stunting in early life. Maternal & child nutrition, 7, 5-18.

[10] Wanjiru, W. H. (2021). Improved balanced random survival forest for the analysis of right censored data: application in determining under five child mortality (Doctoral dissertation, Moi University).

[11] Takele, K., Zewotir, T., & Ndanguza, D. (2019). Understanding correlates of child stunting in Ethiopia using generalized linear mixed models. BMC Public Health, 19(1), 1-8.

[12] Hougaard, P. (1995). Frailty models for survival data. Lifetime data analysis, 1(3), 255-273.

[13] Corsi, D. J., Neuman, M., Finlay, J. E., & Subramanian, S. V. (2012). Demographic and health surveys: a profile. International journal of epidemiology, 41(6), 1602-1613.

[14] Keino, S., Plasqui, G., Ettyang, G., & van den Borne, B. (2014). Determinants of stunting and overweight among young children and adolescents in sub-Saharan Africa. Food and nutrition bulletin, 35(2), 167-178.

[15] UNICEF. (2016). The under-five mortality rate: The indispensable gauge of child health. www.unicef.org/sowc08/docs/sowc08_panels.pdf Accessed Mar 2016.

[16] Black, R. E., Levin, C., Walker, N., Chou, D., Liu, L., Temmerman, M., & Group, D. R. A. (2016). Reproductive, maternal, newborn, and child health: key messages from disease control priorities 3rd edition. The Lancet, 388(10061), 2811-2824.

[17] McGuire, J. W. (2006). Basic health care provision and under-5 mortality: a cross-national study of developing countries. World Development, 34(3), 405-425.

[18] Child Mortality. (2022, January 20). UNICEF DATA. https://data.unicef.org/topic/child-survival/under-five-mortality/

[19] Nations, U. (2016). The Sustainable Development Goals 2016. eSocialSciences. http://www.un.org/sustainabledevelopment/health/

[20] Wanjiru, W. H. (2021). Improved balanced random survival forest for the analysis of right censored data: application in determining under five child mortality (Doctoral dissertation, Moi University).

[21] Bereka, S. G., Habtewold, F. G., & Nebi, T. D. (2017). Under-five mortality of children and its determinants in Ethiopian Somali regional state, Eastern Ethiopia. Health Science Journal, 11(3), 1.

[22] Khan, J. R., & Awan, N. (2017). A comprehensive analysis on child mortality and its determinants in Bangladesh using frailty models. Archives of Public Health, 75(1), 1-10.

[23] Sharma, A. K., & Dutta, R. (2020). Determinants of child survival at the household level: An insight of the method of factor analysis. In Contemporary Issues in Sustainable Development (pp. 253-271). Routledge India.

[24] Corsi, D. J., Neuman, M., Finlay, J. E., & Subramanian, S. V. (2012). Demographic and health surveys: a profile. International journal of epidemiology, 41(6), 1602-1613.

[25] Burnham, K. P. (2000). Model selection and multimodel inference. A practical information-theoretic approach.

[26] Nelson Owuor Onyango (2009). On the Linear Mixed Effects Regression (lmer) R Function for Nested Animal Breeding Data. CS-BIGS 4(1): 44-58. http://erepository.uonbi.ac.ke:8080/xmlui/handle/123456789/38574

[27] Searle, S. R., Casella, G., & McCulloch, C. E. (2009). Variance components. John Wiley & Sons.

[28] Patterson, H. D. (1971). Thompson R. Recovery of inter-block information when block sizes are unequal. Biometrika, 58, 545-554.

[29] Duchateau, L., Janssen, P., & Rowlands, J. (1998). Linear mixed models. An introduction with applications in veterinary research. ILRI (aka ILCA and ILRAD).

[30] Ayiko, R., Antai, D., & Kulane, A. (2009). Trends and determinants of under-five mortality in Uganda. East African journal of public health, 6(2), 136-140.

[31] Nasejje, J. B., Mwambi, H. G., & Achia, T. N. (2015). Understanding the determinants of under-five child mortality in Uganda including the estimation of unobserved household and community effects using both frequentist and Bayesian survival analysis approaches. BMC public health, 15(1), 1003.

[32] Sreeramareddy, C. T., Kumar, H. N., & Sathian, B. (2013). Time Trends and Inequalities of Under-Five Mortality in Nepal: A Secondary Data Analysis of Four Demographic and Health Surveys between 1996 and 2011. PLoS ONE, 8(11): e79818. https://doi.org/10.1371/journal.pone.0079818

[33] Verweij, P. J., & van Houwelingen, H. C. (1995). Time-dependent effects of fixed covariates in Cox regression. Biometrics, 1550-1556.

[34] Cox, D. R., & Oakes, D. (1984). Analysis of survival data. Chapman and Hall/CRC.

[35] Lee, E. T., & Wang, J. (2003). Statistical methods for survival data analysis. (vol. 476). John Wiley & Sons.

[36] Ratnaningsih, D. J., Saefuddin, A., Kurnia, A., & Mangku, I. W. (2019, July). Stratified-extended cox model in survival modeling of non-proportional hazard. In IOP Conference Series: Earth and Environmental Science (vol. 299, No. 1, p. 012023). IOP Publishing.

[37] Vaupel, J. W., Manton, K. G., & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography, 16(3), 439-454.

[38] Weathers, B. (2017). Comparision of Survival Curves Between Cox Proportional Hazards, Random Forests, and Conditional Inference Forests in Survival Analysis.

[39] Muriithi, D. M., & Muriithi, D. K. (2015). Determination of infant and child mortality in Kenya using cox-proportional hazard model. American Journal of Theoretical and Applied Statistics, 4(5), 404-413.

[40] Otieno, F., & Omolo, C. (2003). Infant and child mortality. Kenya demographic and health survey, 114-122.