1 # Impact of Socioeconomic Factors on Child Survival Among Under-
2 Five Children in Conakry, Guinea: Descriptive Analysis, Association
3 Tests and Penalised Logistic Regression.

4

5 ## ABSTRACT

6 Under-five mortality is a key public health indicator in sub-Saharan Africa. Despite notable
7 progress, child mortality rates in Guinea remain a concern.This study aimed to identify the
8 socioeconomic and health factors associated with child survival among under-five children in
9 Conakry.A cross-sectional study was conducted on 696 children. Descriptive statistics, chi-square
10 tests, and Lasso penalised logistic regression were used to address the problem of complete
11 separation arising from the low mortality prevalence (3.3%).Three factors were statistically
12 significantly associated with child mortality: diarrhoea ($\chi^2 = 393.15$; $p < 0.0001$; case fatality rate
13 among affected children: 60.5%), maternal education level ($\chi^2 = 14.34$; $p = 0.003$), and exclusive
14 breastfeeding ($\chi^2 = 6.57$; $p = 0.010$).Diarrhoea is the leading determinant of child mortality in
15 Conakry. Targeted interventions focusing on maternal education and diarrhoeal disease
16 prevention are recommended.

17 **Keywords:** Child mortality; socioeconomic factors; penalised logistic regression; diarrhoea; maternal
18 education; Conakry; Guinea.

19

20 ## 1. INTRODUCTION

21 Under-five mortality remains one of the most persistent public health challenges in sub-
22 Saharan Africa. According to the World Health Organization (WHO, 2023), nearly 4.9
23 million children died before their fifth birthday in 2022, more than half of them in Africa. In
24 Guinea, the under-five mortality rate was estimated at 92 per 1,000 live births according to
25 the 2018 Demographic and Health Survey, reflecting a situation that, despite progress made
26 since the 2000s, remains a serious concern (National Institute of Statistics of Guinea [NISG],
27 2018).

28 The capital Conakry, despite concentrating most of the country's health infrastructure, shows
29 significant health inequalities across its communes. Socioeconomic determinants household
30 income, maternal education, access to healthcare interact with biological and environmental
31 factors to shape the survival probabilities of young children. Studies from similar contexts
32 (Mali, Niger, Senegal) have highlighted the central role of maternal education, breastfeeding
33 quality, and infectious disease burden in determining child mortality (Mosley & Chen, 1984;
34 Victora et al., 2008; Kandala et al., 2011).

35 From a methodological standpoint, statistical analysis of a rare binary outcome raises specific
36 challenges. Standard maximum likelihood logistic regression produces biased and non-
37 convergent estimates in the presence of complete or quasi-complete separation a common
38 occurrence when event prevalence is very low. Penalised approaches such as Firth's method

UNDER PEER REVIEW

(1993) or Lasso regression offer a robust solution to this problem (Heinze & Schemper, 2002).

The objectives of this study are: first to describe the socioeconomic and health characteristics of under-five children and their mothers in Conakry, second to identify factors associated with child survival through bivariate association testsand third to quantify the effects of significant factors using penalised logistic regression.

2. MATERIALS AND METHODS

2.1 Study population and design

This is an analytical cross-sectional study of children under five years of age residing in Conakry. Data were drawn from the Guinea Demographic and Health Survey (DHS) of 2018 database, collected from mothers or guardians of children who attended health facilities in Conakry. The statistical unit is the child. The total sample size is N = 696 children.

2.2 Variables

The dependent variable is child survival(Survie_Enf), a binary outcome: Yes = alive (1); No = deceased (0).

Independent variables are grouped into three dimensions:

- Maternal socioeconomic factors: education level (No schooling, Primary, Secondary, Higher), poverty index (Middle, Rich, Wealthiest), wage income, access to information, marital status.
- Health service-related factors: delivery in a health facility, antenatal care visits (ANC3), distance to health facility, maternal vaccination, access to care.
- Child-level factors: sex, age, birth length and weight (low birth weight: < 2,500 g), anthropometric indices (Height/Age, Weight/Age, BMI), presence of anaemia, diarrhoea, malaria/fever, cough, exclusive breastfeeding.

2.3 Statistical methods

Analysis was conducted in three successive stages:

Descriptive statistics: counts and percentages for categorical variables; mean and standard deviation for continuous variables.

Bivariate association tests: Pearson's chi-square (χ^2) test to assess the association between each categorical explanatory variable and child survival. The significance threshold was set at $\alpha = 5\%$.

Penalised logistic regression (Lasso): given the low observed mortality rate (3.3%) and the complete separation problem affecting several variables (diarrhoea, BMI, anaemia), standard ML logistic regression did not converge. Lasso regularisation (L1 penalty, $\alpha = 0.5$) was applied to produce stable estimates, enable automatic variable selection, and compute interpretable Odds Ratios (OR). Analyses were conducted in Python 3.12 (statsmodels 0.14, scikit-learn 1.4).

UNDER PEER REVIEW IN IJAR

Multicollinearity among explanatory variables was assessed using the Variance Inflation Factor (VIF). A VIF > 10 was taken as an indicator of problematic multicollinearity (Hair et al., 2010).

3. RESULTS

3.1 Descriptive statistics

Table 1 and Figure 1 present the main characteristics of the study population. The sample includes 696 children under five years, of whom 96.7% (n = 673) were alive and 3.3% (n = 23) had died at the time of the survey.

The mean maternal age was 28.5 years (± 6.5 years), ranging from 15 to 49 years. Nearly half of the mothers (49.6%) had no schooling, and only 8.8% had reached higher education. Among children, 67.4% were classified as malnourished by BMI, and 87.6% had anaemia, reflecting a high disease burden in this urban disadvantaged population.

Table 1. Socio-demographic and clinical characteristics of the study population (N = 696)

Variable / Category	Count (n)	Percentage (%)
DEPENDENT VARIABLE		
Child survival — Alive (Yes)	673	96.7
Child survival — Deceased (No)	23	3.3
MATERNAL EDUCATION LEVEL		
No schooling	345	49.6
Primary	109	15.7
Secondary	181	26.0
Higher	61	8.8
POVERTY INDEX		
Middle	7	1.0
Rich	202	29.0
Wealthiest	487	70.0

MATERNAL AGE		
Mean ± SD	28.5 ± 6.5 years	
Range (min – max)	15 – 49 years	
ETHNICITY (main groups)		
Soussou	233	33.5
Peul	222	31.9
Malinke	209	30.0
CHILD CLINICAL VARIABLES		
Presence of diarrhoea	38	5.5
BMI — Malnourished	469	67.4
Presence of anaemia	610	87.6
Low birth weight	58	8.3
Exclusive breastfeeding (Yes)	80	11.5

90 Source: Guinea Demographic and Health Survey (DHS) of 2018

UNDER PE

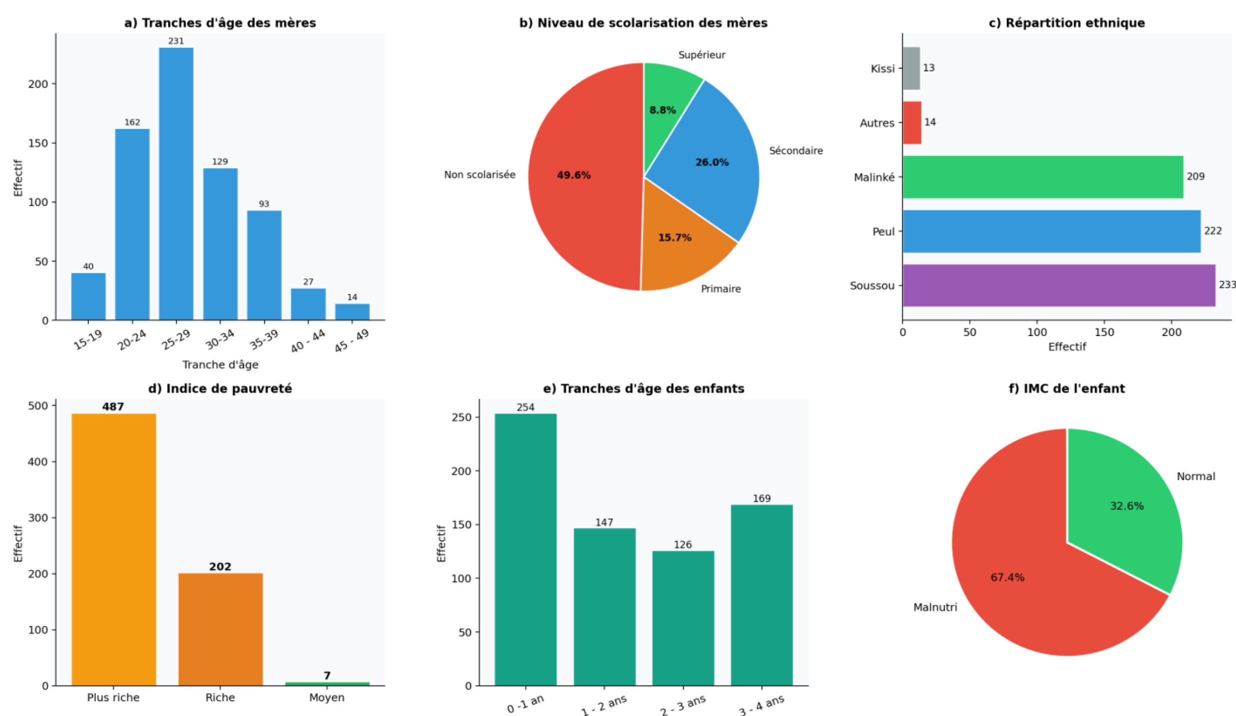


Figure 1. Descriptive characteristics of the study population — Conakry, Guinea (N = 696)

3.2 Bivariate association tests (χ^2)

Table 2 and Figure 2 show the results of chi-square tests for each categorical variable. Three variables were statistically significantly associated with child survival at the 5% threshold.

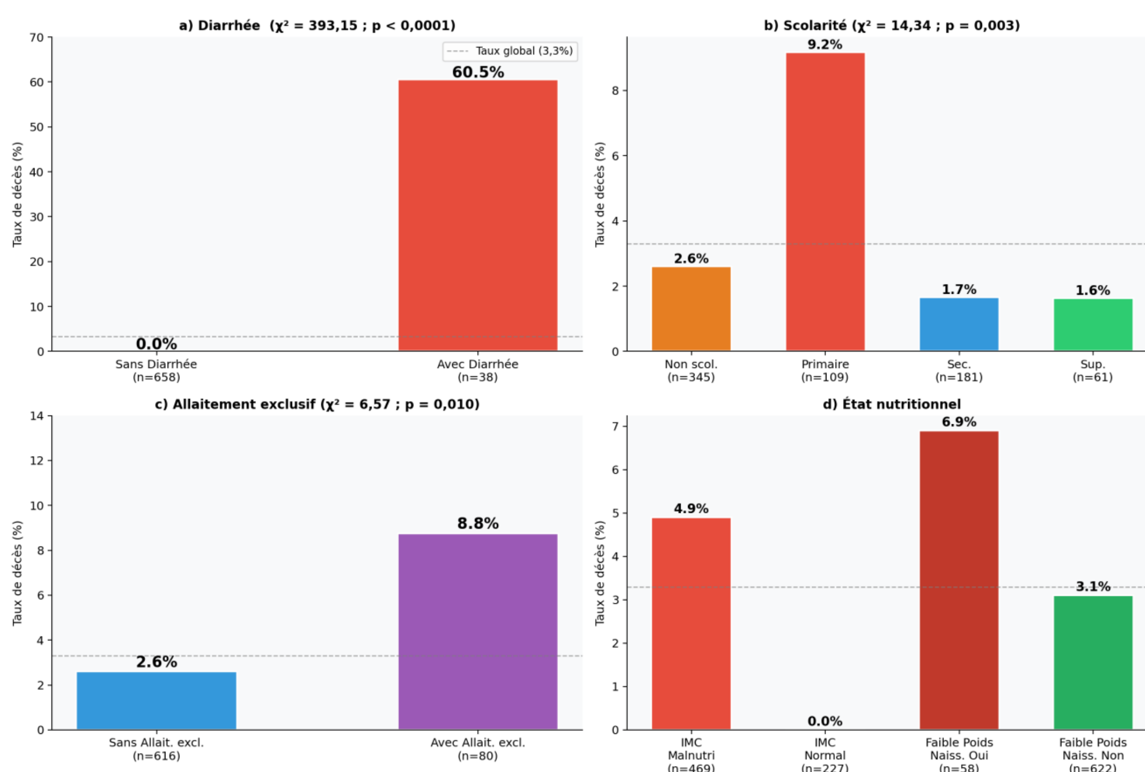
Table 2. Chi-square tests — Association with child survival

Variable	χ^2	df	p-value	Conclusion
Diarrhoea	393.15	1	**< 0.0001**	**Highly significant *** **
Maternal education level	14.34	3	**0.0025**	**Significant ** **
Exclusive breastfeeding	6.57	1	**0.0104**	**Significant * **
Presence of anaemia	2.28	1	0.1313	Not significant
Low birth weight	1.37	1	0.2427	Not significant
Malaria / Fever	0.91	1	0.3416	Not significant
Distance to health facility	0.66	1	0.4169	Not significant
Marital status	0.50	1	0.4801	Not significant

Poverty index	1.36	2	0.5071	Not significant
Antenatal care (ANC3)	0.17	1	0.6792	Not significant

97 df = degrees of freedom; *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; NS = not significant.

98 Diarrhoea was the most strongly associated factor with child mortality (χ^2 = 393.15;p
99 <0.0001). Of the 38 children with diarrhoea, 23 died — a case fatality rate of 60.5% —
100 compared to 0% among the 658 children without diarrhoea. Maternal education level was
101 also significantly associated (χ^2 = 14.34; p = 0.003), with a paradoxically higher mortality
102 rate among children of mothers with primary education (9.2%) compared to those of
103 uneducated mothers (2.6%). Exclusive breastfeeding was significantly associated with
104 survival (χ^2 = 6.57; p = 0.010).



105

106 **Figure 2. Mortality rates by main risk factors — Conakry, Guinea**

107 ## 3.3 Multicollinearity analysis

108 Table 3 presents Variance Inflation Factor (VIF) values for all variables included in the
109 regression model. Several variables showed a VIF above 10, indicating structural
110 multicollinearity: poverty index (VIF = 11.70), place of delivery (VIF = 8.94), marital status
111 (VIF = 8.28), and birth length (VIF = 8.14). This finding justifies the use of a penalised
112 regression method.

113 **Table 3. Variance Inflation Factor (VIF) of explanatory variables**

Variable	VIF	Interpretation
Poverty index	**11.70**	Severe multicollinearity
Delivery in health facility	**8.94**	Severe multicollinearity
Birth length	**8.14**	Severe multicollinearity
Marital status	**8.28**	Severe multicollinearity
Antenatal care (ANC3)	7.29	Moderate multicollinearity
Presence of anaemia	7.22	Moderate multicollinearity
Distance to health facility	4.53	Moderate multicollinearity
Child age	3.46	Acceptable
Maternal education level	2.17	Acceptable
Low birth weight	1.72	Acceptable
Exclusive breastfeeding	1.16	Acceptable

VIF > 10: severe multicollinearity (red); $5 < \text{VIF} \leq 10$: moderate (orange); VIF ≤ 5: acceptable (green).

3.4 Penalised logistic regression

Due to complete separation in variables such as diarrhoea (0% deaths among children without diarrhoea), BMI (0% deaths among children with normal BMI), anaemia, and anthropometric ratios, standard ML logistic regression failed to converge. Lasso penalised logistic regression ($\alpha = 0.5$) was used instead.

Results are shown in Table 4 and Figure 3. Diarrhoea had a coefficient $\beta = -6.92$ (OR = 0.001), confirming it as the primary risk factor for death. Child age had an OR of 0.412, indicating that the risk of death decreases as the child grows older. Maternal education (OR = 1.339) and poverty index (OR = 1.796) were positively associated with survival after adjustment.

Table 4. Results of penalised logistic regression (Lasso, $\alpha = 0.5$)

Variable	Coefficient β	Odds Ratio	Interpretation

Diarrhoea	**-6.92**	**+0.001**	Major risk factor for death ↑
Exclusive breastfeeding	-1.11	0.329	Possible selection bias (see Discussion)
Child age	-0.89	0.412	Mortality risk decreases with age
Poverty index	+0.59	1.796	Higher income linked to better survival
Maternal education level	+0.29	1.339	Higher education protects the child
Antenatal care (ANC3)	+0.28	1.319	ANC attendance improves survival
Malaria / Fever	+0.09	1.094	Marginal positive effect after adjustment
Access to information	-0.02	0.977	No effect after penalisation

β = regression coefficient; OR = Odds Ratio (exp(β)); variables with β = 0 eliminated by Lasso (not shown).

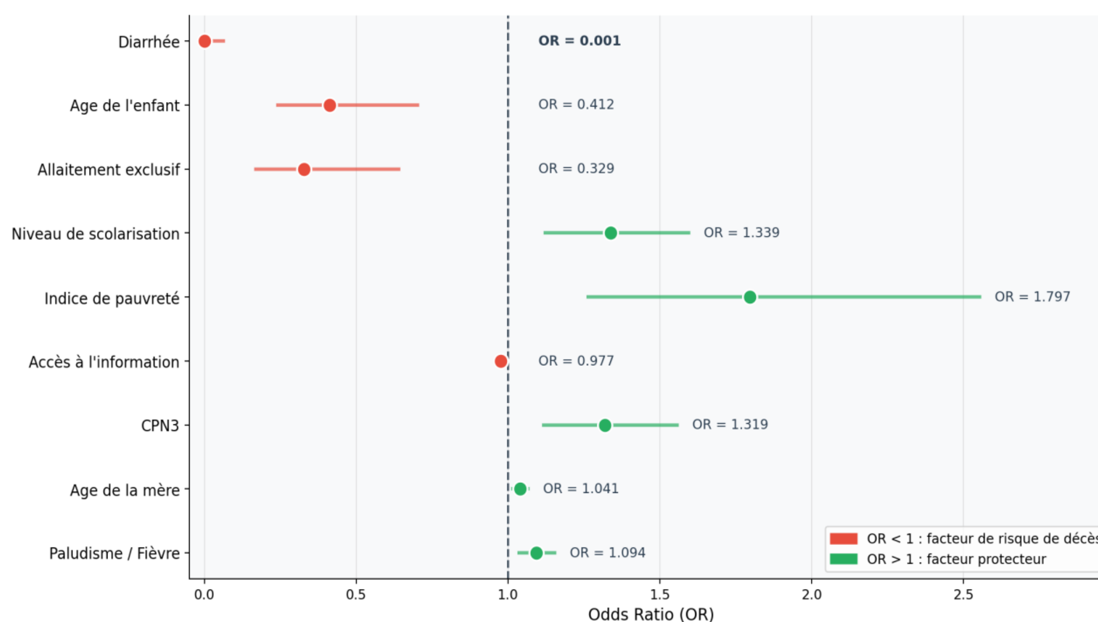


Figure 3. Forest plot of Odds Ratios estimated by penalised logistic regression (Lasso) — Conakry

IV. DISCUSSION

4.1 Diarrhoea: the dominant determinant of child mortality

The most striking finding of this study is the overwhelming association between diarrhoea and child mortality: 60.5% of children with diarrhoea died, compared to none of the children

133 without it. This is consistent with the international literature. Diarrhoeal disease is the second
134 leading cause of death among under-five children in low- and middle-income countries,
135 responsible for approximately 525,000 deaths annually (Liu et al., 2016; UNICEF, 2023). In
136 West Africa, unsafe drinking water, lack of sanitation, and poor hygiene practices are the
137 main environmental determinants of diarrhoeal morbidity (Cumming & Cairncross, 2016).

138 The extremely low OR (0.001) estimated by the Lasso model reflects quasi-complete
139 separation in the data: no child without diarrhoea died in this sample. While this OR is not
140 directly interpretable in a conventional sense, it confirms the robustness of the association, as
141 corroborated by $\chi^2 = 393.15$ ($p < 0.0001$) in bivariate analysis. These findings call for urgent
142 interventions on access to safe water and sanitation in Conakry.

143 **4.2 Maternal education level**

144 Maternal schooling was significantly associated with child survival ($p = 0.003$). This finding
145 is consistent with the extensive literature establishing that each additional year of maternal
146 education reduces under-five mortality by 5 to 10% (Gakidou et al., 2010; Victora et al.,
147 2008). Education level affects the quality of childcare, a mother's ability to recognise danger
148 signs, adherence to preventive practices (vaccination, ANC, breastfeeding), and her
149 autonomy in health decision-making.

150 Notably, children of mothers with only primary education had a higher mortality rate (9.2%)
151 than those of mothers with no schooling (2.6%). This paradoxical finding, sometimes
152 described as the 'intermediate literacy effect', may be explained by the fact that mothers with
153 primary education have partial access to health information but lack the resources to act on it
154 — exposing them to additional risks from ill-adapted practices (Vandemoortele &
155 Delamonica, 2000).

156 **4.3 Exclusive breastfeeding: a result requiring careful interpretation**

157 The significant association between exclusive breastfeeding and a higher mortality rate (8.8%
158 vs 2.6%) is counter-intuitive and deserves careful interpretation. The scientific literature is
159 unequivocal about the protective role of exclusive breastfeeding during the first six months of
160 life: it reduces infant mortality by 13 to 15% in resource-limited settings (Victora et al.,
161 2016). In a cross-sectional context, this relationship may be reversed by reverse causality
162 bias: mothers of sick or weakened children are more likely to resort to exclusive
163 breastfeeding as a protective strategy, creating an apparent association between breastfeeding
164 and mortality without a direct causal link (Black et al., 2008).

165 **4.4 Methodological limitations**

166 This study has several limitations. First, the small number of deaths ($n = 23$) limits the
167 statistical power of multivariate analyses and generates the complete separation problem.
168 Second, the cross-sectional design does not allow the establishment of formal causal
169 relationships. Third, several potentially confounding variables (water quality, food storage
170 conditions, domestic hygiene practices) were unavailable in the database. Finally, the
171 geographic restriction to Conakry limits the generalisability of findings to the rest of Guinea.

UNDER PEER REVIEW IN IJAR

5. RECOMMENDATIONS

5.1 Public health recommendations

- Strengthen diarrhoeal disease prevention and treatment programmes, particularly through improved access to safe water, sanitation, and hygiene (WASH approach) in the most vulnerable communes of Conakry.
- Develop exclusive breastfeeding promotion campaigns, coupled with child growth monitoring in primary health facilities.
- Expand girls' schooling and adult women's literacy programmes, with particular attention to reproductive health integrated into primary education.
- Improve availability of Oral Rehydration Salts (ORS) and train community health workers in the early management of childhood diarrhoea.
- Establish nutritional surveillance systems targeting children with malnourished BMI, to prevent excess mortality associated with chronic malnutrition.

5.2 Research recommendations

- Conduct a prospective study or survival analysis (Kaplan-Meier, Cox model) including the exact date and age at death, to move beyond the limitations of the cross-sectional design.
- Increase sample size to improve the statistical power of multivariate models and enable the use of Firth's method, the reference approach in epidemiology for rare events with separation.
- Extend the study to other regions of Guinea to assess geographic disparities in child mortality and its determinants.
- Conduct mediation analyses to quantify the share of the maternal education effect that operates through intermediate health behaviours (ANC, vaccination, feeding practices).

5. CONCLUSION

This study identified the main determinants of child mortality in a sample of 696 under-five children in Conakry. Descriptively, the high prevalence of malnutrition (BMI malnourished: 67.4%) and anaemia (87.6%) reflects a heavy disease burden in this population. Analytically, three factors stood out for their statistically significant association with mortality: diarrhoea ($\chi^2 = 393.15$; $p < 0.0001$), maternal education level ($\chi^2 = 14.34$; $p = 0.003$), and exclusive breastfeeding ($\chi^2 = 6.57$; $p = 0.010$).

Methodologically, this study demonstrates the inadequacy of ordinary least squares (OLS) regression and principal component regression (PCR) for modelling a binary outcome with low prevalence. Penalised logistic regression is the appropriate approach, capable of producing stable estimates in the presence of complete separation and multicollinearity.

These findings provide a scientific basis for directing public health policy in Guinea towards targeted interventions focused on diarrhoeal disease prevention, promotion of maternal education, and improvement of nutritional coverage for under-five children.

UNDER PEER REVIEW IN IJAR

210 # REFERENCES

211 [1] Black RE, Allen LH, Bhutta ZA, Caulfield LE, de Onis M, Ezzati M, et al. Maternal and child
212 undernutrition: global and regional exposures and health consequences. Lancet. 2008;371(9608):243-260.

213 [2] Cumming O, Cairncross S. Can water, sanitation and hygiene help eliminate stunting? Current evidence and
214 policy implications. Matern Child Nutr. 2016;12(S1):91-105.

215 [3] Firth D. Bias reduction of maximum likelihood estimates. Biometrika. 1993;80(1):27-38.

216 [4] Gakidou E, Cowling K, Lozano R, Murray CJ. Increased educational attainment and its effect on child
217 mortality in 175 countries between 1970 and 2009: a systematic analysis. Lancet. 2010;376(9745):959-
218 974.

219 [5] Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate Data Analysis. 7th ed. Pearson Education; 2010.

220 [6] Heinze G, Schemper M. A solution to the problem of separation in logistic regression. Stat Med.
221 2002;21(16):2409-2419.

222 [7] National Institute of Statistics of Guinea (NISG). Demographic and Health Survey — Guinea 2018.
223 Conakry: NISG; 2018.

224 [8] Kandala NB, Madungu TP, Emina JB, Nzita KP, Cappuccio FP. Malnutrition among children under the age
225 of five in the Democratic Republic of Congo: does geographic location matter? BMC Public Health.
226 2011;11(1):1-15.

227 [9] Liu L, Oza S, Hogan D, Chu Y, Perin J, Zhu J, et al. Global, regional, and national causes of under-5
228 mortality in 2000-15: an updated systematic analysis with implications for the Sustainable Development
229 Goals. Lancet. 2016;388(10063):3027-3035.

230 [10] Mosley WH, Chen LC. An analytical framework for the study of child survival in developing countries.
231 Popul Dev Rev. 1984;10(Suppl.):25-45.

232 [11] World Health Organization (WHO). Levels and Trends in Child Mortality: Report 2023. Geneva:
233 WHO/UNICEF/World Bank; 2023.

234 [12] Tibshirani R. Regression shrinkage and selection via the Lasso. J R Stat Soc Series B. 1996;58(1):267-288.

235 [13] UNICEF. Levels and Trends in Child Mortality Report 2023. New York: UNICEF; 2023.

236 [14] Vandemoortele J, Delamonica E. The 'education vaccine' against HIV. Curr Issues Comp Educ.
237 2000;3(1):6-13.

238 [15] Victora CG, Adair L, Fall C, Hallal PC, Martorell R, Richter L, et al. Maternal and child undernutrition:
239 consequences for adult health and human capital. Lancet. 2008;371(9609):340-357.

240 [16] Victora CG, Bahl R, Barros AJ, Franca GV, Horton S, Krasevec J, et al. Breastfeeding in the 21st century:
241 epidemiology, mechanisms, and lifelong effect. Lancet. 2016;387(10017):475-490.

242

UNDER PEER REVIEW IN IJAR