



The Adolescent Attendance Gap: A Survey-Weighted Analysis of School Attendance among Children in Kogi State, Nigeria (MICS6, 2021)

¹Dolapo Shobanke, ¹Grace Korter, ²Praise Lawal Molete and ^{*}Elijah Babatunde Sunday

¹Department of Statistics, Federal University, Lokoja, Kogi State, Nigeria.

²Department of Statistics, Kwara State University, Malete, Kwara State, Nigeria.

*Corresponding authors' email: elisbabs499@gmail.com

ABSTRACT

Nigeria has one of the largest populations of out-of-school children in the world, but subnational, state-level evidence on the correlates of school attendance remains limited, particularly for north-central states such as Kogi. This study estimated the prevalence of current school attendance among children aged 5–17 years in Kogi State and examined its sociodemographic correlates using the 2021 Nigeria Multiple Indicator Cluster Survey (MICS6). The analytic sample comprised 583 children nested within 49 primary sampling units (PSUs). A null random-intercept model assessing PSU-level clustering was singular (variance not estimable); survey-weighted logistic regression accounting for stratification, clustering, and weights was therefore the primary analysis, with survey-weighted Poisson regression yielding adjusted prevalence ratios (APRs) as a sensitivity analysis. The weighted prevalence of current school attendance was 95.0% (95% CI: 92.0–96.9%), declining from 98.8% among children aged 5–9 years to 88.2% among adolescents aged 15–17 years. In adjusted models, no covariate was statistically significant at the 5% level, including female sex (adjusted odds ratio [AOR] 0.84, 95% CI 0.25–2.81), age 12–17 years (AOR 0.32, 95% CI 0.07–1.34), and rural residence (AOR 0.71, 95% CI 0.22–2.35); APRs were consistent. The wide confidence intervals are compatible with both negligible and substantial associations, so these results do not demonstrate an absence of sociodemographic differences; only 28 children were not attending school, limiting statistical power. Attendance was high overall, with the largest shortfall among older adolescents; larger or pooled samples are needed to characterise its correlates, and retention measures for older adolescents merit consideration and evaluation.

Received: 10 Sep. 2026

Accepted: 13 Oct. 2026

Published: 14 Oct. 2026

Keywords:

School Attendance, Out-Of-School Children, MICS6, Survey-Weighted Regression, Adolescents

INTRODUCTION

Education is a foundational determinant of child development and a recognised driver of long-term economic and social wellbeing. Sustainable Development Goal 4 (SDG 4) commits countries to ensuring inclusive and equitable quality education for all children by 2030, yet Nigeria continues to report the largest absolute number of out-of-school children of any country globally, with recent estimates from the UNICEF global database placing roughly one in three Nigerian children outside the formal schooling system (UNICEF Data, 2025; National Bureau of Statistics [NBS] & United Nations Children's Fund [UNICEF], 2022; Nwoke et al., 2024).

The 2021 Nigeria Multiple Indicator Cluster Survey (MICS6), implemented by the National Bureau of Statistics with technical support from UNICEF, documented marked geographic disparities in school participation, with the highest out-of-school rates concentrated in north-western and north-eastern states and comparatively lower rates in southern and parts of north-central Nigeria (NBS & UNICEF, 2022; UNICEF Nigeria, 2023; Adeleke, 2024). Kogi State lies in the North-Central geopolitical zone, and the MICS6 education fact sheets report out-of-school rates for the state of 6% at primary-school age, 5% at lower-secondary age, and 15% at upper-secondary age, compared with national rates of 26%, 25%, and 34%, respectively (UNICEF Nigeria, 2023). These figures are among the lowest of any state in the three northern

zones. Whether this favourable position reflects state education policy or other contextual factors cannot be established from cross-sectional survey data, and attributing it to specific investments or fee-support programmes would require dedicated evaluation. However, favourable aggregate figures can mask residual inequities by age, sex, household wealth, maternal education, and place of residence that are only visible through disaggregated, model-based analysis. The same state-level figures show that out-of-school rates in Kogi are higher at upper-secondary age than at primary or lower-secondary age, which motivates a closer look at age differences within the state.

Across sub-Saharan Africa and Nigeria more broadly, established correlates of non-attendance include older age and the transition from primary to secondary schooling, female sex (particularly in northern contexts), rural residence, low household wealth, and low parental education (UNICEF Nigeria, 2023; Education Policy and Data Center, 2013; Ojuolape & Mohd, 2024; Pezzulo et al., 2022). Adolescents aged 12–17 years are consistently at elevated risk of dropout as direct and indirect schooling costs rise, early marriage becomes more prevalent, and competing demands for adolescent labour intensify. These patterns make it important to assess whether the same drivers operate in a relatively high-attendance state such as Kogi, or whether the high overall attendance masks a more specific vulnerability concentrated among older age groups.

Because school attendance may be correlated within communities, multilevel (random-effects) models are often considered for cluster-sampled household surveys. Where non-attendance is rare, however, such models can fail to converge or return singular variance estimates (Bates et al., 2015), and survey-weighted regression with design-based standard errors is a practical alternative (Lumley, 2010). Because attendance was also expected to be highly prevalent, odds ratios can diverge from prevalence ratios, so a survey-weighted Poisson model with robust variance was specified as a sensitivity analysis (Zou, 2004). The statistical rationale is set out in Section 2.4.

The aim of this study was to estimate the survey-weighted prevalence of current school attendance among children aged 5–17 years in Kogi State using the 2021 MICS6 and to examine its sociodemographic correlates while accounting for the complex survey design. The title refers to the adolescent attendance gap because the clearest descriptive difference in this population occurs among older adolescents (15–17 years), although the analytic sample covers the full 5–17 age range.

MATERIALS AND METHODS

Study Design and Data Source

This study is a secondary, cross-sectional analysis of the Kogi State sample of the 2021 Nigeria Multiple Indicator Cluster Survey (MICS6), a nationally representative household survey implemented by the National Bureau of Statistics (NBS) with technical and financial support from UNICEF and partner agencies (NBS & UNICEF, 2022). MICS6 used a stratified, two-stage cluster sampling design in which enumeration areas were selected with probability proportional to size (PPS) within each state stratum, followed by systematic random selection of households within each selected enumeration area. Fieldwork was conducted between mid-2021 and December 2021, with the 2020–2021 school year serving as the reference period for education questions. Household members were listed, and modules relevant to child school participation were completed by knowledgeable adult respondents.

Data were drawn from the Nigeria MICS6 SPSS datasets, using the child-level file (fs.sav) as the primary source and the household-level file (hh.sav) merged by the household identifiers HH1 and HH2. The sample was restricted to children residing in Kogi State (HH7 = 22), aged 5–17 years, with valid information on current school attendance. The variables used were CB7 (current school attendance), CB3 (age), HL4 (sex), HH6 (residence), melevel (maternal education), windex5 (wealth index quintile), fsweight (sampling weight), stratum (sampling stratum), and PSU (primary sampling unit). The sampling weight fsweight was used as supplied, without rescaling or normalisation, and the survey design was specified with PSU as the cluster identifier, stratum as the stratification variable, and fsweight as the weight, with clusters nested within strata.

Study Population and Sample

The analytic population comprised children aged 5–17 years residing in surveyed households in Kogi State, for whom school attendance and all covariates included in the adjusted models were recorded. The final analytic sample comprised 583 children, of whom 555 were currently attending school and 28 were not, nested within 49 PSUs across Kogi State.

Variables

The outcome variable was current school attendance, defined from the survey item on whether the child attended school or

an early childhood programme during the 2020–2021 school year, coded as a binary indicator (1 = attending, 0 = not attending). Candidate covariates included child sex (male/female), child age (grouped into 5–9, 10–14, and 15–17 years), place of residence (urban/rural), household wealth derived from the MICS combined wealth index (poorest, second, middle, fourth, richest quintiles), and maternal or caretaker educational attainment (none, primary, junior secondary, senior secondary, higher/tertiary). Household size (1–4, 5–7, and 8 or more members) was also tabulated descriptively in Table 1 but was not included in the adjusted models.

Because only 28 children in the analytic sample were not attending school, the full set of covariate categories could not be supported in a single adjusted model without producing unstable estimates. Prior to model fitting, age was collapsed into two groups (5–11 years and 12–17 years), maternal education into two groups (none/primary versus secondary or higher), and household wealth into three groups (poorer [poorest and second quintiles combined], middle, and richer [fourth and richest quintiles combined]).

Statistical Analysis

Survey-Weighted Prevalence Estimation

All analyses accounted for the MICS6 complex sample design, incorporating sampling weights w_i , stratification, and PSU-level clustering through survey-design objects in R (Lumley, 2010). The design-weighted ratio estimator of the weighted prevalence of school attendance, with Taylor series linearisation for variance estimation, is:

$$\bar{Y}_w = \frac{\sum_i w_i Y_i}{\sum_i w_i}, \quad \text{Prev}(\%) = 100 \times \bar{Y}_w \quad (1)$$

where $\bar{Y}_w$ is the survey-weighted proportion of children currently attending school, $\text{Prev}(\%)$ is the same quantity expressed as a percentage, $Y_i \in \{0, 1\}$ is the binary attendance outcome for child i (1 = attending, 0 = not attending), and w_i is the survey-assigned sampling weight; the sums run over all children in the analytic sample. The 95% confidence interval was constructed using Taylor series linearisation to account for stratification and PSU clustering.

Multilevel Feasibility Assessment

To assess whether attendance outcomes clustered meaningfully by PSU, a null (intercept-only) random-intercept logistic regression model was fitted, with children at level 1 nested within PSUs at level 2. The model is:

$$\text{logit}(p_{ij}) = \gamma_0 + u_j, \quad u_j \sim N(0, \sigma_u^2) \quad (2)$$

Where p_{ij} is the probability that child i in PSU j is currently attending school, $\text{logit}(p) = \ln[p/(1-p)]$, γ_0 is the overall intercept (fixed effect) and u_j is the PSU-specific random intercept, assumed to follow $N(0, \sigma_u^2)$, with PSUs assumed independent. Fitting this model yields an estimate σ_u^2 of the between-PSU variance. A singular fit is flagged when this estimate lies on the boundary of the parameter space:

$$\hat{\sigma}_u^2 \rightarrow 0 \Rightarrow \text{singular fit (boundary estimate)} \quad (3)$$

A singular fit indicates that the optimizer reached the boundary of the parameter space, so that the between-PSU variance could not be estimated as different from zero in this sample. It should be read as a limitation of estimability rather than as proof that community-level variation is absent: with very few non-attending children spread across 49 PSUs, the data carry little information about between-PSU variance, and a boundary estimate can arise even when some true heterogeneity exists. Singular fits are a recognised feature of mixed models fitted to sparse data (Bates et al., 2015). Related guidance on small numbers of grouping levels (Oberpriller et al., 2022; Gomes, 2022) concerns situations

with very few clusters, which is not the case here (49 PSUs), so the singularity is better attributed to the scarcity of non-attending children.

The survey-weighted approach addresses clustering differently from a multilevel model. A random-intercept model represents between-PSU heterogeneity explicitly as a model component and targets cluster-specific (conditional) effects, whereas design-based estimation treats PSU clustering as a feature of the sampling process, adjusts the standard errors for it through the sandwich variance in equation (6), and targets population-average (marginal) effects. The design-based standard errors remain valid whether or not a random-intercept variance can be estimated, which is why this approach was used as the primary analysis; it does not, however, show that clustering is absent.

For completeness, the intraclass correlation coefficient (ICC) would ordinarily quantify the proportion of total variance in the outcome propensity attributable to between-PSU differences, using the logistic latent-variable approximation:

$$ICC = \frac{\sigma_u^2}{\sigma_u^2 + \pi^2/3} \quad (4)$$

Where σ_u^2 is the between-PSU variance and $\pi^2/3 \approx 3.29$ is the fixed level-1 residual variance on the logit scale. When the variance is estimated at zero, the ICC is zero and the median odds ratio (MOR) equals 1 by construction; these values reflect the boundary estimate rather than a demonstrated absence of clustering. Because the variance was not estimable, ICC and MOR were not reported as substantive findings.

Primary Analysis: Survey-Weighted Logistic Regression

Because the between-PSU variance could not be estimated in the null multilevel model, and because design-based estimation accommodates the complex sample design directly, survey-weighted logistic regression was used as the primary adjusted analytic approach. Binary logistic regression is widely used to model sociodemographic correlates of binary outcomes (Idikwu et al., 2025; Usman et al., 2021). Let $p_{ij} = \Pr(Y_{ij} = 1)$ denote the probability of current attendance for child i in PSU j . The model is specified as:

$$\text{logit}(p_{ij}) = \ln\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_0 + \beta_1 X_{1ij} + \dots + \beta_k X_{kij} \quad (5)$$

Where X_{1ij} , X_{kij} are the collapsed covariates described in Section 2.3 (child sex, age group, residence, maternal education group, wealth group), and β_0 , β_1 , β_k are the regression coefficients. The model is estimated by pseudo-maximum likelihood, with each observation weighted by w_{ij} . Rather than a model-based (random-effects) covariance estimator, the variance of β is obtained using the design-based Taylor series linearisation sandwich estimator:

$$\hat{V}(\hat{\beta}) = (X^T \hat{W} X)^{-1} \hat{V}_{\text{Taylor}}(X^T \hat{W} X)^{-1} \quad (6)$$

where X is the design matrix of the covariates, $\hat{W}$ is the working weight matrix incorporating w_{ij} and the logistic variance function, and $\hat{V}_{\text{Taylor}}$ is the Taylor series linearisation estimate of the sampling variance of the weighted score contributions, which captures the stratified-cluster contribution to sampling variance through PSU-level score sums and the between-PSU variability in weighted residuals. This estimator is designed to account for design effects due to stratification and clustering without requiring a correctly specified multilevel variance structure (Lumley, 2010). The exponentiated coefficient, $\exp(\hat{\beta}_k)$, gives the adjusted odds ratio (AOR):

$$AOR = \exp(\hat{\beta}_k) \quad (7)$$

Interpreted as the multiplicative change in the odds of current school attendance associated with a one-unit or one-category change in covariate k , adjusted for all other covariates in the

model. A two-sided p-value below 0.05 was taken as the threshold for statistical significance.

Sensitivity Analysis: Survey-Weighted Poisson Regression

Because current school attendance was highly prevalent (95.0%), odds ratios can overstate the magnitude of association relative to the prevalence-ratio scale when the outcome is common. Following Zou (2004), a survey-weighted Poisson regression model with robust variance was fitted as a sensitivity analysis to estimate adjusted prevalence ratios (APRs) directly, an approach whose properties relative to alternative robust-variance estimators have since been examined further (Talbot et al., 2023). The model is:

$$\ln(\mu_{ij}) = \alpha_0 + \alpha_1 X_{1ij} + \dots + \alpha_k X_{kij} \quad (8)$$

Where $\mu_{ij} = E(Y_{ij}) = \Pr(Y_{ij} = 1)$ is the prevalence of attendance. Coefficients α_0 , α_1 , α_k are estimated by pseudo-maximum likelihood under the Poisson log-likelihood, weighted by w_{ij} . Although the Poisson variance assumption is misspecified for binary data, this does not introduce bias into the coefficient estimates; it only affects the standard errors. The variance is therefore corrected using the cluster-robust sandwich estimator:

$$\hat{V}_{\text{robust}}(\hat{\alpha}) = (X^T \hat{\Lambda} X)^{-1} (\sum_c X_c^T \hat{e}_c \hat{e}_c^T X_c) (X^T \hat{\Lambda} X)^{-1} \quad (9)$$

where $\hat{\Lambda}$ is a diagonal matrix of fitted Poisson means, X_c is the covariate matrix for cluster (PSU) c , $\hat{e}_c = y_c - \mu_c$ is the residual vector for cluster c , summed over all clusters to propagate the design-based contribution to variance; in the survey-weighted fit, the sampling weights are incorporated in these score contributions. The exponentiated coefficient gives the adjusted prevalence ratio (APR):

$$APR = \exp(\hat{\alpha}_k) \quad (10)$$

Interpreted as the multiplicative change in the prevalence of current school attendance associated with a unit change in covariate k , all else equal. An APR below 1 indicates a lower prevalence of current school attendance than in the reference group, and an APR above 1 indicates a higher prevalence. Because the outcome is attendance rather than an adverse event, APR values are not described as protective or risk effects. Unlike the AOR, which overstates the prevalence ratio when the outcome is common, the APR is directly interpretable as a ratio of attendance prevalences.

Model Diagnostics and Sparse-Data Considerations

Only 28 children were not attending school, and six non-reference coefficients were estimated in addition to the intercept, giving about 4.7 events per estimated coefficient, well below the roughly 10 events per variable often cited as a guide for logistic regression (Peduzzi et al., 1996). Estimates from such sparse data can be unstable, and odds ratios and their confidence intervals can be biased or imprecise; the wide intervals in Table 3 are consistent with this concern. Categories were therefore collapsed before modelling (Section 2.3), and results are interpreted with explicit attention to imprecision. No formal diagnostics for multicollinearity, for the influence of individual PSUs or strata, or for the goodness of fit of the final survey-weighted logistic model were performed. The only model check implemented was the assessment of singularity in the null random-intercept model using the `isSingular()` function in `lme4`, with the `bobyqa` optimiser and a maximum of 2×10^5 function evaluations; this limitation is acknowledged in Section 3.5. A penalised (Firth-type) logistic regression (Firth, 1993) is a natural sensitivity analysis for sparse data; it was not part of the present analysis and is identified as a priority for future work with larger samples.

Ethical Considerations

This study used de-identified, publicly available secondary survey data. The original MICS6 survey obtained ethical clearance and informed consent from respondents at the time of data collection; no additional ethical approval was required for this secondary analysis. All analyses were conducted using R with the survey package (Lumley, 2010) and lme4 package (Bates et al., 2015). All analyses were conducted in R version 4.5.2, using the survey package (version 4.5) for complex survey analyses and lme4 (version 2.0-1) for the null random-intercept logistic model. Data management and presentation additionally used haven (2.5.5), dplyr (1.1.4), broom (1.0.11), broom.mixed (0.2.9.6), flextable (0.9.10), ggplot2 (4.0.3), and forcats (1.0.1).

RESULTS AND DISCUSSION

Sample Characteristics and Overall Prevalence

The analytic sample included 583 children aged 5–17 years (49.8% male, 50.2% female, weighted), drawn from 49 PSUs

across Kogi State. Of these, 555 (95.2% unweighted) were currently attending school and 28 were not. The overall survey-weighted prevalence of current school attendance was 95.0% (95% CI: 92.0–96.9%).

Table 1 presents the descriptive characteristics and subgroup-specific weighted attendance prevalence for the analytic sample. The most pronounced subgroup variation was observed by age: attendance prevalence declined steadily from 98.8% (95% CI: 96.6–99.6%) among children aged 5–9 years to 96.1% (95% CI: 87.7–98.8%) among those aged 10–14 years, and further to 88.2% (95% CI: 79.5–93.5%) among adolescents aged 15–17 years. Attendance was broadly similar across sex (95.4% boys, 94.6% girls), residence (96.4% urban, 94.2% rural), maternal education groups (92.9% to 100.0%), and wealth quintiles (92.0% to 97.8%), with wide and overlapping confidence intervals across most subgroups, reflecting the small number of non-attending children in each category.

Table 1: Descriptive Characteristics and School Attendance Prevalence among Children Aged 5–17 Years in Kogi State, Nigeria (MICS6, 2021)

Variable	Category	Unweighted n	Weighted %	Attendance (%)	95% CI
Child sex	Male	285	49.8	95.4	91.4–97.6
	Female	298	50.2	94.6	88.5–97.5
Age group	5–9 years	230	39.2	98.8	96.6–99.6
	10–14 years	195	33.1	96.1	87.7–98.8
	15–17 years ^a	158	27.8	88.2	79.5–93.5
Residence	Urban	174	35.8	96.4	89.8–98.8
	Rural	409	64.2	94.2	90.4–96.5
Maternal education	None	173	27.7	94.9	86.8–98.1
	Primary	162	30.0	92.9	86.0–96.5
	Junior secondary ^b	20	3.9	100.0	100–100
	Senior secondary	154	25.9	95.4	86.5–98.5
Wealth quintile	Higher/tertiary	74	12.6	97.8	92.6–99.4
	Poorest	66	9.9	97.8	92.0–99.4
	Second	142	22.5	92.0	83.4–96.4
	Middle	180	30.9	94.4	86.0–97.9
	Fourth	130	24.0	97.3	91.9–99.1
Household size	Richest	65	12.6	95.0	83.2–98.6
	1–4 members	207	19.5	94.2	88.5–97.2
	5–7 members	277	45.9	96.7	93.6–98.3
Total	8+ members	99	34.7	93.2	84.3–97.2
		583	100.0	95.0	92.0–96.9

Note. Frequencies are unweighted. Weighted percentages and school-attendance prevalence estimates account for MICS survey weights, stratification, and PSU clustering.

^a Subgroup with the most pronounced decline in attendance prevalence; requires targeted monitoring.

^b Small cell frequency ($n = 20$ unweighted)

Multilevel Feasibility Assessment

The null random-intercept logistic model, fitted to determine whether meaningful between-PSU variation in school attendance existed before covariate adjustment, produced a singular fit, with an estimated PSU-level random-intercept variance of $\sigma_u^2 = 0.000000$ (Table 2). This outcome indicates that, within this analytic sample of 583 children across 49 PSUs, the optimizer reached the boundary of the parameter space, so that between-PSU variance in the log-odds of attendance could not be estimated as different from zero.

This finding is consistent with two features of the data, both of which are plausible but neither of which can be confirmed here: first, the small number of non-attending children ($n = 28$) spread across 49 PSUs provides limited statistical

information with which to estimate between-cluster variance; second, Kogi State's relatively uniform and high attendance levels may leave little community-level variation for a random intercept to capture. Consistent with recommended practice (Bates et al., 2015), intraclass correlation (ICC), median odds ratio (MOR), and proportional change in variance (PCV) were not computed or reported as substantive estimates from a degenerate model. Survey-weighted regression, which still accounts for stratification and clustering through design-based standard errors, was therefore adopted as the primary analytic strategy; this decision rests on the non-estimability of the variance component and not on evidence that community-level clustering is absent.

Table 2: Multilevel Feasibility Assessment for Current School Attendance

Item	Value
Analytic sample (n)	583
Number of PSUs	49
Non-attending children (n)	28
PSU-level random-intercept variance (σ^2_u)	0.000000
Singular fit	Yes
ICC, MOR, PCV reported?	No (not meaningfully estimable)
Analytic decision	Survey-weighted regression as primary approach

Note. The null random-intercept logistic model produced a singular fit, indicating that the PSU-level variance estimate lay on the boundary (zero) in this sample. This reflects limited estimability in a sparse-event setting and is not evidence that PSU-level variation is absent. Multilevel indices (ICC, MOR, PCV) are therefore not reported.

Adjusted Correlates of Current School Attendance

In the survey-weighted logistic regression model adjusting for child sex, age group, residence, maternal education group, and household wealth group, none of the six non-reference covariate categories was statistically associated with current school attendance at the 5% significance level (Table 3). The most substantively notable point estimate was for older adolescents aged 12–17 years relative to children aged 5–11 years (AOR 0.32, 95% CI 0.07–1.34, $p = 0.115$), consistent with the descriptive prevalence decline at older ages, but the confidence interval was wide and crossed unity. Point estimates indicated slightly lower odds of attendance among

girls than boys (AOR 0.84, 95% CI 0.25–2.81, $p = 0.779$) and among rural than urban children (AOR 0.71, 95% CI 0.22–2.35, $p = 0.568$), while the estimates for household wealth did not follow a consistent gradient relative to poorer households (middle: AOR 0.94, 95% CI 0.18–5.02, $p = 0.944$; richer: AOR 1.37, 95% CI 0.29–6.52, $p = 0.689$). Secondary or higher maternal education was also associated with somewhat higher odds (AOR 1.36, 95% CI 0.31–5.91, $p = 0.679$). All of these non-significant results should be interpreted cautiously, as the adjusted models had very limited statistical power given only 28 non-attending children, and a non-significant estimate should not be read as evidence of no association.

Table 3: Survey-Weighted Adjusted Odds Ratios for Current School Attendance, Kogi State

Variable	Category	AOR	95% CI	p-value
Child sex	Male	Ref.	—	—
	Female	0.84	0.25–2.81	0.779
Age group	5–11 years	Ref.	—	—
	12–17 years	0.32	0.07–1.34	0.115
Residence	Urban	Ref.	—	—
	Rural	0.71	0.22–2.35	0.568
Maternal education	None/primary	Ref.	—	—
	Secondary/higher	1.36	0.31–5.91	0.679
Wealth group	Poorer	Ref.	—	—
	Middle	0.94	0.18–5.02	0.944
	Richer	1.37	0.29–6.52	0.689

Note. Adjusted odds ratios estimated using survey-weighted logistic regression. Reference categories: male sex, age 5–11 years, urban residence, none/primary maternal education, poorer households. Confidence intervals were estimated using Taylor series linearisation to account for the stratified, clustered survey design.

Figure 1 presents the adjusted odds ratios and their 95% confidence intervals on a logarithmic scale. All confidence intervals include the null value of 1.0, which is consistent with the non-significant results in Table 3; the intervals are wide and are compatible with both no association and

associations of substantial magnitude. The widest intervals on the log scale are observed for the wealth categories, followed by older age and maternal education, which is consistent with the small number of non-attending children in each category.

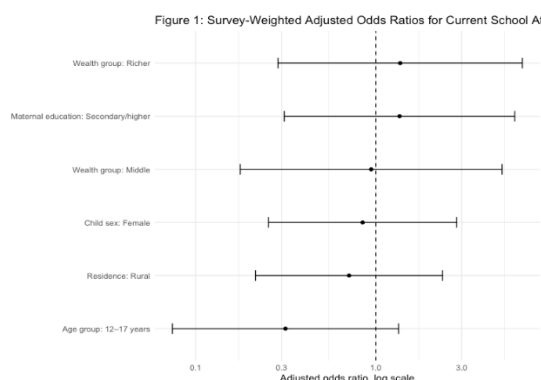


Figure 1: Survey-weighted adjusted odds ratios for current school attendance, Kogi State (log scale). All confidence intervals include 1.0; none of the associations was statistically significant at the 5% level

Sensitivity Analysis: Adjusted Prevalence Ratios

Because the prevalence of current school attendance was 95.0% — a scenario in which odds ratios can substantially overstate the magnitude of prevalence-ratio-scale associations — a survey-weighted Poisson regression model with robust variance was additionally fitted to estimate adjusted prevalence ratios (APRs) for the same set of covariates. Results are presented in Table 4. Consistent with the primary odds-ratio model, no covariate was statistically associated with attendance, and all APR point estimates were close to 1.0. The estimates were directionally consistent with the logistic model: APR for female sex was 0.99 (95% CI 0.93–1.05, $p = 0.755$); APR for age 12–17 years was 0.95 (95% CI 0.89–1.01, $p = 0.087$); APR for rural residence was 0.98 (95% CI 0.93–1.04, $p = 0.567$); APR for secondary or

higher maternal education was 1.01 (95% CI 0.95–1.08, $p = 0.696$); APR for middle wealth was 1.00 (95% CI 0.91–1.10, $p = 0.959$); and APR for richer wealth was 1.01 (95% CI 0.94–1.10, $p = 0.713$). The confidence intervals are numerically narrower on the prevalence-ratio scale than on the odds-ratio scale. Because attendance prevalence is close to 1, prevalence ratios lie closer to 1.0 than odds ratios, so their intervals are correspondingly narrower in appearance; this should not be read as greater precision or additional information. Both analyses use the same 583 children, the same 28 non-attending children, and the same covariates. The APR results are therefore consistent with the primary analysis and suggest that the absence of statistical significance did not depend on the choice of effect measure, but they do not independently confirm it.

Table 4: Survey-Weighted Adjusted Prevalence Ratios for Current School Attendance (Sensitivity Analysis)

Variable	Category	APR	95% CI	p-value
Child sex	Male	Ref.	—	—
	Female	0.99	0.93–1.05	0.755
Age group	5–11 years	Ref.	—	—
	12–17 years	0.95	0.89–1.01	0.087
Residence	Urban	Ref.	—	—
	Rural	0.98	0.93–1.04	0.567
Maternal education	None/primary	Ref.	—	—
	Secondary/higher	1.01	0.95–1.08	0.696
Wealth group	Poorer	Ref.	—	—
	Middle	1.00	0.91–1.10	0.959
	Richer	1.01	0.94–1.10	0.713

Note. Adjusted prevalence ratios estimated using survey-weighted Poisson regression with robust (sandwich) variance, implemented to account for the high prevalence of the outcome (95.0%). Reference categories as in Table 3.

Discussion

This analysis of the Kogi State sample of MICS6 found that current school attendance among children aged 5–17 years was high overall at 95.0%, exceeding several national and northern-region attendance benchmarks reported for Nigeria. The 2021 MICS6 national report documented pronounced disparities with north-eastern and north-western states reporting substantially lower attendance, highlighting Kogi State's comparatively favourable position (NBS & UNICEF, 2022; Oyekan et al., 2023). This pattern is consistent with the MICS-EAGLE state profiles, which place Kogi among the states with the lowest out-of-school rates in the northern zones, with 6% of primary-school-age, 5% of lower-secondary-age, and 15% of upper-secondary-age children out of school (UNICEF Nigeria, 2023). Our estimate that 11.8% of 15–17-year-olds were not attending school is of a similar order to the upper-secondary figure, although the two are not directly comparable because they differ in definition, age band, and analytic sample. Whether state education policy contributed to these favourable figures cannot be determined from these data; any attribution to specific budget allocations or fee-support programmes would require evaluation designs capable of supporting causal inference.

The most policy-relevant finding from the descriptive analysis is the pronounced age gradient: attendance prevalence fell from 98.8% among 5–9-year-olds to 88.2% among 15–17-year-olds, a pattern mirroring well-documented national and regional trends in which attrition accelerates at the transition from primary to secondary education. Across Nigeria, this transition is associated with rising direct and indirect costs, early marriage for girls, and growing labour demands for boys, factors that intersect and compound at the household level (UNICEF Nigeria, 2023;

UNICEF Data, 2025). Even in a high-attendance state, an attendance prevalence of 88.2% among older adolescents implies that roughly one in nine children in this age group is not in school, representing a non-trivial burden that merits attention. This age gradient is a descriptive, unadjusted difference: the 95% confidence intervals for children aged 5–9 and 15–17 years do not overlap, but this is not a formal test. The adjusted analysis, which compared ages 12–17 with 5–11 years, did not provide statistically significant evidence of an age association (AOR 0.32, $p = 0.115$; APR 0.95, $p = 0.087$). Because that contrast pools ages 12–14, a band in which attendance was still high in the descriptive data (96.1% for ages 10–14), with ages 15–17, it may dilute the gap seen in the oldest group, and the descriptive and adjusted results should not be read as contradicting one another.

The preliminary assessment of multilevel structure was a useful step. Although school-attendance outcomes are frequently assumed to cluster by community due to shared school access, infrastructure quality, and local norms, the null random-intercept model returned a singular fit with an estimated between-PSU variance of zero. This should be understood as a limitation of estimability rather than evidence that community-level variation is absent. Two explanations are plausible: with 28 non-attending children spread across 49 PSUs, the data contain little information from which to identify between-cluster variance, and Kogi State's high and fairly uniform attendance may leave little community-level variation for a random intercept to capture. These data cannot establish which explanation holds, and both may operate. Rather than forcing a degenerate multilevel model, this study adopted survey-weighted regression, which addresses design effects through robust standard errors as advocated by Lumley (2010). This choice keeps the standard errors valid

for the complex design, but it does not resolve whether PSU-level heterogeneity exists.

In the adjusted models, no covariate reached statistical significance, and the absence of significant findings should not be interpreted as evidence against the existence of underlying disparities. With only 28 events, the models were severely underpowered: for example, the adjusted odds ratio for older adolescents was 0.32, a substantively notable point estimate, but the confidence interval spanned 0.07 to 1.34, reflecting extreme uncertainty. Point estimates were directionally consistent with known patterns (lower odds for older adolescents, girls, and rural residents, and higher odds for the richer wealth group, although the pattern for wealth was not monotonic; Ojuolape & Mohd, 2024; Pezzulo et al., 2022) but could not be distinguished from chance variation. In neighbouring Kwara State, Ojuolape and Mohd (2024) identified age and educational costs as significant determinants of out-of-school status using household-head survey data and multinomial models; differences in design, outcome definition, and sample size mean that their results and ours are not directly comparable, but both are consistent with age being a salient correlate. National MICS6 profiles likewise show higher out-of-school rates among rural households, poorer households, and children whose mothers have no education (UNICEF Nigeria, 2023), which is the direction of our rural-residence and maternal-education point estimates. The prevalence-ratio sensitivity analysis using Zou's (2004) modified Poisson approach produced directionally similar estimates on the prevalence-ratio scale, with numerically narrower intervals that reflect the different scale rather than additional information. Because both analyses use the same small sample and covariates, this consistency shows only that the non-significant findings were not specific to one effect measure; it does not confirm that they are correct. Taken together, these findings are best interpreted as consistent with but not confirmatory of the literature on educational marginalisation in Nigeria, and they underscore the need for larger-sample or pooled-survey analyses for this state.

This study has several limitations. First, only 28 children were not attending school, so the adjusted models had low power, produced wide intervals, and may be affected by sparse-data bias; a penalised analysis was not undertaken and should be considered in future work, noting that implementations combining penalisation with design weights are less established. Second, formal diagnostics for multicollinearity, for the influence of individual PSUs or strata, and for the goodness of fit of the final model were not performed. Third, categories were collapsed to stabilise estimation, which reduces the detail available on, for example, wealth and maternal education gradients. Fourth, attendance was reported by knowledgeable adult household respondents, and the design is cross-sectional and based on a single survey round, so the findings describe associations at one time point and cannot support causal conclusions. Finally, between-PSU variance could not be estimated, so possible community-level heterogeneity in attendance remains unexamined.

This observational analysis did not evaluate any intervention, so its findings cannot show that particular programmes would raise attendance in Kogi State. They do point to older adolescents as the group in which a shortfall is most evident. External evidence from other low- and middle-income settings indicates that cash transfers, whether conditional or unconditional, improve the odds of school enrolment and attendance (Baird et al., 2014), and fee waivers and scholarships have also been proposed as responses to cost barriers at secondary level, although this study provides no

evidence on their effects in Kogi State. Whether such approaches would be effective, affordable, or cost-effective there is an empirical question that would need to be assessed through dedicated evaluation.

CONCLUSION

Current school attendance among children aged 5–17 years in Kogi State, Nigeria, was high (95.0%) in this 2021 MICS6 sample, with the most pronounced shortfall concentrated among adolescents aged 15–17 years (88.2%). In the multilevel feasibility check, PSU-level variance could not be estimated as different from zero (a singular fit); this reflects limited estimability in a sparse-event setting and does not show that community-level clustering is absent. Survey-weighted logistic regression with design-based standard errors was therefore used as the primary adjusted analysis, and survey-weighted Poisson regression was used as a sensitivity analysis for adjusted prevalence ratios. Adjusted associations between sex, age, residence, maternal education, and household wealth and current attendance were all statistically non-significant; given only 28 non-attending children, this is better read as a reflection of limited statistical power than as evidence that these factors do not matter.

These findings support continued targeted attention to adolescent retention in Kogi State's education system, particularly at the primary-to-secondary transition point and among older teenagers. Measures that address the direct and indirect costs of secondary schooling, such as fee waivers, conditional cash transfers, and scholarship programmes, are possibilities informed by external evidence (Baird et al., 2014); this study did not test them, and they should be regarded as options for consideration and further evaluation in Kogi State rather than demonstrated solutions. Larger-sample or pooled-survey analyses combining multiple MICS rounds, or linking MICS data with administrative school records, would help to more precisely characterise the sociodemographic correlates of non-attendance in this relatively high-attendance context.

REFERENCES

- Adeleke, R. (2024). Geographical analysis of gender disparity in out-of-school children in Nigeria. *Child Indicators Research*, 17, 637–655. <https://doi.org/10.1007/s12187-023-10097-z>
- Baird, S., Ferreira, F. H. G., Özler, B., & Woolcock, M. (2014). Conditional, unconditional and everything in between: A systematic review of the effects of cash transfer programmes on schooling outcomes. *Journal of Development Effectiveness*, 6(1), 1–43. <https://doi.org/10.1080/19439342.2014.890362>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Education Policy and Data Center. (2013). *Nigeria: Out-of-school children profile*. FHI 360. https://www.epdc.org/sites/default/files/documents/Nigeria_OOSC_Profile.pdf
- Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80(1), 27–38. <https://doi.org/10.1093/biomet/80.1.27>

- Gomes, D. G. E. (2022). Should I use fixed effects or random effects when I have fewer than five levels of a grouping factor in a mixed-effects model? *PeerJ*, 10, Article e12794. <https://doi.org/10.7717/peerj.12794>
- Idikwu, J.S., Kuhe, D.A., & Nyitar, J. A. (2025). A binary logistic regression model for assessing the predisposing factors of malaria fever among farmers in Benue State, Nigeria. *FUDMA Journal of Sciences*, 9(12), 22-28. <https://doi.org/10.33003/fjs-2025-0912-4147>
- Lumley, T. (2010). *Complex surveys: A guide to analysis using R*. Wiley.
- National Bureau of Statistics & United Nations Children's Fund. (2022). *Nigeria Multiple Indicator Cluster Survey 2021, survey findings report*.
- Nwoke, C., Oyiga, S., & Cochrane, L. (2024). Assessing the phenomenon of out-of-school children in Nigeria: Issues, gaps and recommendations. *Review of Education*, 12(3), Article e70011. <https://doi.org/10.1002/rev3.70011>
- Oberpriller, J., de Souza Leite, M., & Pichler, M. (2022). Fixed or random? On the reliability of mixed-effects models for a small number of levels in grouping variables. *Ecology and Evolution*, 12(7), Article e9062. <https://doi.org/10.1002/ece3.9062>
- Ojuolape, A. M., & Mohd, S. (2024). Determinants of child schooling in Kwara State, North Central Nigeria. *Cogent Social Sciences*, 10(1), Article 2319377. <https://doi.org/10.1080/23311886.2024.2319377>
- Oyekan, K., Ayorinde, A., & Adenuga, O. (2023). *The problem of out-of-school children in Nigeria* (RISE Insight Note 2023/058). Research on Improving Systems of Education (RISE) Programme. https://doi.org/10.35489/BSG-RISE-RI_2023/058
- Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, 49(12), 1373–1379. [https://doi.org/10.1016/S0895-4356\(96\)00236-3](https://doi.org/10.1016/S0895-4356(96)00236-3)
- Pezzulo, C., Alegana, V. A., Christensen, A., Bakari, O., & Tatem, A. J. (2022). Understanding factors associated with attending secondary school in Tanzania using household survey data. *PLOS ONE*, 17(2), Article e0263734. <https://doi.org/10.1371/journal.pone.0263734>
- Talbot, D., Mésidor, M., Chiu, Y., Simard, M., & Sirois, C. (2023). An alternative perspective on the robust Poisson method for estimating risk or prevalence ratios. *Epidemiology*, 34(1), 1–7. <https://doi.org/10.1097/EDE.0000000000001544>
- UNICEF Data. (2025). *Primary school age education*. <https://data.unicef.org/topic/education/primary-education/>
- UNICEF Nigeria. (2023). *Nigeria education fact sheets: Analyses for learning and equity using MICS data (MICS-EAGLE)*. <https://www.unicef.org/nigeria/reports/nigeria-education-fact-sheets>
- Usman, U., Waziri, M., Manu, F., Zakari, Y., & Dikko, H.G. (2021). Assessing the performance of some re-sampling methods using logistic regression. *FUDMA Journal of Sciences*, 5(1), 445-456. <https://doi.org/10.33003/fjs-2021-0501-589>
- Zou, G. (2004). A modified Poisson regression approach to prospective studies with binary data. *American Journal of Epidemiology*, 159(7), 702–706. <https://doi.org/10.1093/aje/kwh090>

