

Urbanization And HIV/AIDS Mortality In Tamale Metropolis, Ghana: A Logistic Regression Model

¹Fuseini Kassim, ²Salifu Katara, and ³Shei Baba Sayibu

^{1,2,3}*Faculty of Physical Sciences, Department of Statistics,
University for Development Studies, Tamale, Ghana, West Africa
Corresponding Author: sheibaba@uds.edu.gh*

Abstract

The study set out to build a predictive model identifying which urbanization-related factors raise HIV/AIDS mortality risk in Tamale Metropolis, Ghana, a city where 80% of residents live in urban areas and HIV prevalence (0.68%) is more than double the regional average (0.31%). A mixed-methods design was used with 300 urban residents. Quantitative data came from structured questionnaires; qualitative data came from 12 interviews and 4 focus group discussions. The model selection was based on deviance, AIC, BIC, and likelihood ratio tests. Firth's penalized regression corrected separation problems. The best model (M_8) had an AUC of 0.864, 96.4% sensitivity, and 52.0% specificity. Six significant predictors emerged: housing type (OR = 20.40), duration without services (OR = 8.63), income (OR = 6.35), distance to health facilities (OR = 5.78), employment (OR = 3.33), and ART access (OR = 0.15, protective). Qualitative data explained the statistics: informal housing enables infection and poor outcomes; distance reflects flooded roads and lost wages; ART access is undermined by stigma. Counterintuitive findings (income and employment increasing risk) were explained by informal, mobile work patterns. Structural factors such as housing, healthcare access, and socioeconomic status are powerful drivers of HIV/AIDS mortality, often exceeding biomedical factors. The model is a practical screening tool, and urban planning must integrate health considerations to achieve SDGs 3 and 11.

Keywords: Urbanization, mortality risk, predictive modelling, regimen, structural determinants

1. Introduction

Urbanization is a transformative force reshaping societies worldwide, particularly in sub-Saharan Africa (Sulemana, 2019). In Ghana, the urban population has been steadily increasing, with cities experiencing significant growth (GSS, 2021). Tamale Metropolis, the capital of the Northern Region, is the third-largest urban area in

Ghana, with approximately 80.8% of its 200,000 residents living in urban settings (Fuseini, 2016). This study investigates the critical link between rapid urbanization and HIV/AIDS mortality in the Tamale Metropolis of Ghana, aiming to develop a predictive model to inform better public health and urban planning interventions. The city faces an HIV prevalence rate of 0.68%, as indicated in the 2019 Ghana AIDS Commission report, which is higher than the regional average of 0.31% (Abdul-Samed et al., 2024). Furthermore, structural factors such as inequality, gender disparity, and lack of infrastructure can compound these challenges (Akinwale, 2023; Marie and Igbal, 2024; Olawade, 2024; Argaw et al., 2024). Factors such as increased population density, mobility, and socioeconomic challenges prevalent in urban settings can exacerbate the spread of HIV (Ullah et al., 2021). Rapid urbanization in Tamale Metropolis brings both opportunities and challenges. On one hand, urban centres can offer improved access to healthcare, education, and employment. On the other hand, they often face issues such as overcrowding, inadequate infrastructure, and heightened health risks. One significant health concern in urban Ghana is the prevalence of HIV/AIDS. Nationally, approximately 274,600 people are living with HIV, with an adult prevalence rate of 2.4% (Nutor, 2021). The impact of HIV/AIDS on mortality in Northern Ghana is significant. Studies have identified HIV/AIDS as one of the leading communicable diseases contributing to mortality in the region (Dwomoh, 2018). For instance, a review of medical certificates of cause of death at the Tamale Teaching Hospital revealed that HIV/AIDS accounted for 12.4% of communicable disease-related deaths (Dwomoh, 2018). Despite the availability of antiretroviral therapy (ART) and efforts to expand its coverage, challenges remain. While ART coverage in Ghana increased from 1% in 2004 to 60% in 2020, gaps in access to treatment and adherence persist. Barriers such as stigma, discrimination, and limited healthcare infrastructure hinder the effectiveness of HIV/AIDS interventions (Sefah et al., 2022).

The review synthesizes global and local research, establishing the complex relationship between urbanization and health. While urbanization can improve healthcare access and life expectancy, it is also linked to environmental degradation, inequality, and specific health risks. Studies in African cities show that slum dwellers and rural-urban migrants often exhibit riskier sexual behaviours. However, mortality can decline with improved access to care (e.g., a 32% decline in Nairobi slums). Understanding how urbanization contributes to HIV/AIDS-related mortality is crucial for developing targeted interventions (Mensah, 2022). Despite national progress in reducing HIV prevalence, public health strategies often fail to integrate urban-specific factors such as housing density, healthcare accessibility, and social dynamics into their frameworks. This oversight leaves critical gaps in understanding how urbanization amplifies the mortality risks associated with HIV/AIDS (Marzukhi, 2023). Addressing this gap requires a holistic approach that combines urbanization metrics with HIV/AIDS prevalence to model mortality risks effectively, ultimately supporting evidence-based public health interventions (Musonda et al., 2023). HIV testing rates are still low among sexually active unmarried young men in sub-Saharan Africa. Based on these findings, health policymakers may need to revisit existing HIV prevention policies and programs to redesign current testing campaigns to increase

uptake among young people (Musonda et al., 2023). Cultural practices also carry many myths surrounding HIV/AIDS, particularly in relation to ritual initiation and traditional circumcision. In many rural communities, cultural norms have undermined women and put them at greater risk of infection. Living conditions have also contributed significantly to the spread of HIV/AIDS, yet unfortunately, environmental, cultural, and social factors receive limited attention in efforts to address the epidemic (Linganiso and Gwegweni, 2016).

The study will examine population density, healthcare access, housing quality, and socioeconomic status as key urbanization variables. The study seeks to address questions about the key predictors of high HIV/AIDS mortality risk associated with rapid urbanization in Tamale Metropolis, and how these can be modelled to inform public health interventions and urban planning.

Current public health strategies fail to integrate these urban-specific factors into HIV/AIDS interventions. There is a lack of localized models to understand how urbanization amplifies mortality risk. The findings will inform evidence-based interventions, guide urban planning, reduce stigma, and support Sustainable Development Goals 3 (Good Health) and 11 (Sustainable Cities). The insights generated could guide multi-sectoral collaborations and ensure that urban development strategies in Tamale Metropolis and beyond prioritize health and well-being for all (Kusturica, 2023). The main aim is to model the mortality risk associated with urbanization and HIV/AIDS. The rest of the article is organized as follows. Section 2 describes the study area and design, while Section 3 covers data collection methods. Logistic regression is undertaken in Section 4. Section 5 deals with Firth's penalized likelihood logistic regression, and the conclusion and recommendations are covered in Section 6.

Study Area and Design

Tamale Metropolis, the capital of Ghana's Northern Region, is a key economic, administrative, and cultural hub with approximately 200,000 residents, 80% living in urban areas. Located about 600 km north of Accra, it shares borders with Sagnarigu Municipal District. Rapid urbanization has created dense settlements and informal housing with overcrowded health facilities and uneven healthcare access. The city's HIV prevalence of 0.68% (representing about 2,137 PLHIV), more than the regional average of 0.31%, reflects urban health burdens. The predominantly Dagomba and Muslim population (GSS, 2021) presents cultural norms and stigma affecting HIV/AIDS responses.

Tamale Metropolis was chosen for its rapid urbanization and disproportionately high HIV prevalence compared to the surrounding region. Urban residents aged 18 years and above, with a focus on vulnerable groups (informal settlement dwellers, PLHIV, women, low-income residents). A disproportionate sampling technique was used to ensure representation across socioeconomic and geographic lines.

2. Methodology and Data Collection

The primary population included urban residents exposed to urbanization challenges, with special attention to informal settlement dwellers, individuals living with HIV/AIDS, women, and low-income residents. Ethical clearance was sought through

the Ghana Health Service and from individuals in the selected sample. Structured questionnaires collected quantitative data on demographics, socioeconomic status, healthcare access, living conditions, and HIV/AIDS awareness/ART access. Qualitative data were gathered through semi-structured interviews and focus group discussions with healthcare providers, policymakers, community leaders, and individuals living with HIV/AIDS as presented in Section 4.4. Statistical analyses were conducted using JASP 0.95.4.0 software. A disproportionate sampling technique ensured representation across socioeconomic, demographic, and geographic characteristics based on gender, income levels, and residential types (formal/informal settlements). Sample size was determined using Cochran's formula (1977):

$$n = \frac{Z^2 P(1-P)}{e^2} \quad (1)$$

Where $Z=1.96$ (95% confidence), $p=0.5$ (maximum variability), and $e = 0.05$ (5% margin of error).

$$n_0 = \frac{1.96^2 \times 0.5 \times 0.5}{0.05^2} = 384.16 \approx 384.$$

Adjusting for the finite population (since 1360 is finite),

$$n = \frac{n_0}{1 + \frac{n_0 - 1}{N}} = \frac{384}{1 + \frac{384}{1360}} \approx 299.6.$$

Hence, 300 respondents in total for population-level representation were selected. The selection was done using systematic random sampling with a sampling interval, $k = \frac{1360}{300} = 4$.

3.1. Logistic Regression Model Tree

Used to model binary high HIV/AIDS mortality risk ($y = 1$) against low HIV/AIDS mortality risk ($y = 0$) based on self-reported perceptions as predictors:

$$\ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n + \quad (2)$$

Where p = probability of high mortality risk, and $X_1, X_2, \dots, X_n$ represent independent variables (population density, healthcare prevalence, etc.). The variables were coded as shown in Table 1.

Table1: Variables categorization

Dependent Variable		
Variables	Label	Categories
	High Urbanization health risk	0 = no 1 = yes
Independent Variables		
Variables	Label	Categories
Age	Age distribution of Respondents	0 = below 45 1 = 45+
Gender	Gender of Respondents	0 = Female 1 = male
Education	Educational Status of Respondents	0 = no formal education 1 = formal education
Employment	Employment Status of Respondents	0 = unemployed 1 = employed
Income	Average monthly income of Respondents	1 = below GHC500 0 = at least GHC500
Duration	Duration of Stay in Tamale of Respondents	0 = below 10 years 1 = At least 10 years
Housing	Housing Type of Respondents	0 = informal 1 = formal
Water	Potable Water Access	0 = no 1 = yes
Distance	Distance from the respondents' residence to the nearest accredited health facility	0 = below 1km 1 = at least 1km
Hchallenges	Any Health Challenges	0 = no 1 = yes
Awareness	Awareness of HIV	0 = no 1 = yes
Tested	Ever Tested HIV	0 = no 1 = yes
ART	Access to ART	0 = no 1 = yes
Sigma	Stigma exist	0 = no 1 = yes

Source: Constructed by the author using primary data.

4. Logistic Regression Model Formulation

Table 2 compares eleven logistic regression models (M_0 to M_{10}), each adding a new predictor variable to determine which model best explains the data. Deviance and model fitness (AIC, BIC) were employed to measure the performance of the model. Lower values generally indicate a better balance between goodness-of-fit and model complexity. There is a consistent decrease in Deviance from 336.8 to 223.5 and AIC from 338.825 to 245.484 from M_0 to M_{10} . This suggests that each additional variable improves the model's fit, with the most complex model being M_{10} . However, it should be noted that the BIC penalizes complexity more heavily. Increases from M_9 onwards (from 283.406 to 286.189), suggesting that the improvements from M_9 to M_{10} might be marginal when accounting for the additional parameters.

The Chi-Square Change (ΔX^2) and p-values test whether the addition of a new variable significantly improves the model compared to the previous one. The change in M_1 (Distance) is significant ($p < 0.01$), meaning adding "Distance" significantly improved the model over the intercept-only model (M_0). The improvement in M_2 (Duration) is significant ($p < 0.01$). This suggests that adding "Duration" is a statistically robust improvement to the model. The improvement in M_3 (Housing) is also significant ($p = 0.001$). M_4 , which is "Hchallenge," is also statistically significant ($p = 0.002$). However, Education inclusion in M_5 is not significant ($p = 0.988$), which means it does not improve the performance of the model. The addition of ART in M_6 is also significant ($p < 0.01$), indicating it provides a meaningful improvement to the model. The inclusion of income in M_7 is also significant ($p = 0.002$). The last two steps (M_9 and M_{10}) show p-values of .068 and .088 for the ΔX^2 . This means that the addition of potable water access in M_9 and Age in M_{10} did not significantly improve the model beyond M_8 . From a parsimony standpoint (keeping the model simple), M_8 is likely the best final model, as the additional complexity of M_9 and M_{10} is not statistically justified, as shown in Table 2.

Furthermore, it can also be deduced from Table 3 that the Pseudo R^2 values (McFadden, Nagelkerke, Tjur) improved consistently from M_0 to M_{10} . Model M_8 reports the Nagelkerke R^2 as 0.445 and Tjur R^2 as 0.356, which are considered moderate-to-strong effect sizes. Moreover, observing M_9 and M_{10} shows that, while the R^2 increases slightly in these models, the addition of potable water access and Age caused a massive inflation of the standard errors for the Intercept and other variables. Overall, the Model Summary suggests that model M_8 has the best explanatory power; the incremental benefit of adding variables after M_8 is not important, as the p-values for ΔX^2 are often not significant.

Table 3: Model Summary for High HIV/AIDS Mortality Risk

Mo del	Devi ance	AIC	BIC	df	ΔX_2	p	McFa dden R ²	Nagel kerke R ²	Tj ur R ²	Co x & Sn ell R ²
M ₀	336. 8	338. 825	342. 525	2 9 8			0.000		0.0 00	
M ₁	313. 4	317. 404	324. 805	2 9 7	23. 421	<. 00 1	0.070	0.111	0.0 86	0.0 75
M ₂	297. 2	303. 194	314. 295	2 9 6	16. 210	<. 00 1	0.118	0.184	0.1 37	0.1 24
M ₃	285. 1	293. 059	307. 861	2 9 5	12. 135	<. 00 1	0.154	0.235	0.1 72	0.1 59
M ₄	273. 1	283. 077	301. 579	2 9 4	11. 983	<. 00 1	0.189	0.284	0.2 14	0.1 92
M ₅	262. 4	274. 448	296. 651	2 9 3	10. 628	.00 1	0.221	0.326	0.2 30	0.2 20
M ₆	246. 5	260. 545	286. 448	2 9 2	15. 903	<. 00 1	0.268	0.386	0.2 94	0.2 61
M ₇	236. 5	252. 520	282. 124	2 9 1	10. 025	.00 2	0.298	0.422	0.3 41	0.2 85
M ₈	229. 7	247. 729	281. 033	2 9 0	6.7 92	.00 9	0.318	0.445	0.3 56	0.3 01
M ₉	226. 4	246. 401	283. 406	2 8 9	3.3 27	.06 8	0.328	0.457	0.3 67	0.3 09
M ₁₀	223. 5	245. 484	286. 189	2 8 8	2.9 18	.08 8	0.336	0.467	0.3 71	0.3 15

Table 4 presents the model coefficients. Distance (OR = 5.78) shows that a one-unit increase in distance multiplies the odds of high HIV/AIDS mortality risk by 5.78. This is the strongest continuous predictor. Duration (OR = 8.63) indicates the highest odds ratio in the core model. Housing (OR = 20.40) is the most powerful predictor. The odds of high HIV/AIDS mortality Risk are over 20 times higher for this specific housing type compared to the baseline. Health challenges show an OR = 0.85. This variable strongly increases the odds of high HIV/AIDS mortality risk. For every unit increase in this variable, the odds decrease by 15%. Access to ART indicates an OR = 0.15. Access to ART is associated with 85% lower odds of high mortality risk. The income variable shows OR = 6.35, whereas Employment has OR = 3.33. These values may be a result of other unmeasured confounding factors such as the type of employment (regular or irregular) from which a respondent gets his or her income. The variable "Education" consistently has a massive negative coefficient (approximately -18) and a standard error in the thousands, resulting in p-values near 1.00 (such as 0.986).

Table 4: Models' Coefficients

Model	Parameter	Estimate	Standard Error	Odds Ratio	z	Wald Test		
						Wald Statistic	df	p
M ₀	(Intercept)	1.094	0.133	2.987	8.202	67.267	1	<.001
M ₁	(Intercept)	-1.479	0.535	0.228	-2.765	7.644	1	.006
	Distance	1.479	0.304	4.386	4.863	23.651	1	<.001
M ₂	(Intercept)	-4.045	0.877	0.018	-4.612	21.269	1	<.001
	Distance	1.760	0.329	5.815	5.349	28.614	1	<.001
	Duration	1.245	0.313	3.475	3.983	15.868	1	<.001
M ₃	(Intercept)	-6.779	1.243	0.001	-5.454	29.747	1	<.001
	Distance	1.723	0.331	5.603	5.2	27.175	1	<.001

Table 4: Models' Coefficients

Model	Parameter	Estimate	Standard Error	Odds Ratio	z	Wald Test		
						Wald Statistic	df	p
M ₄					13			.01
	Duration	1.869	0.374	6.481	4.993	24.928	1	<.001
	Housing	1.463	0.448	4.317	3.264	10.656	1	.001
	(Intercept)	-4.381	1.452	0.013	-3.016	9.096	1	.003
	Distance	1.585	0.338	4.880	4.691	22.003	1	<.001
	Duration	1.988	0.389	7.300	5.113	26.148	1	<.001
M ₅	Housing	1.518	0.463	4.563	3.280	10.760	1	.001
	Hchallenge	-1.350	0.429	0.259	-3.147	9.903	1	.002
	(Intercept)	13.301	1139.574	597513.696	0.012	1.362×10 ⁻⁴	1	.991
	Distance	1.350	0.340	3.856	3.971	15.768	1	<.001
	Duration	1.767	0.386	5.853	4.582	20.992	1	<.001
	Housing	1.547	0.451	4.695	3.430	11.762	1	<.001
M ₆	Hchallenge	-1.393	0.431	0.248	-3.231	10.438	1	.001
	Education	-16.988	1139.573	4.189×10 ⁻⁸	-0.015	2.222×10 ⁻⁴	1	.988
	(Intercept)	14.251	1125.796	1.545×10 ⁺⁶	0.013	1.602×10 ⁻⁴	1	.990

Table 4: Models' Coefficients

Model	Parameter	Estimate	Standard Error	Odds Ratio	z	Wald Test		
						Wald Statistic	df	p
M ₇	Distance	1.624	0.371	5.073	4.374	19.130	1	<.001
	Duration	1.894	0.416	6.643	4.552	20.719	1	<.001
	Housing	2.291	0.529	9.887	4.327	18.724	1	<.001
	Hchallenge	-1.940	0.487	0.144	-3.986	15.889	1	<.001
	Education	-17.614	1125.795	2.240×10 ⁻⁸	-0.016	2.448×10 ⁻⁴	1	.988
	ART	-1.460	0.387	0.232	-3.776	14.259	1	<.001
	(Intercept)	11.704	1090.810	121064.292	0.011	1.151×10 ⁻⁴	1	.991
	Distance	1.856	0.397	6.398	4.670	21.805	1	<.001
	Duration	1.854	0.427	6.388	4.341	18.847	1	<.001
	Housing	2.889	0.586	17.975	4.929	24.291	1	<.001
	Hchallenge	-1.882	0.479	0.852	-3.932	15.457	1	<.001
	Education	-18.149	1090.808	1.312×10 ⁻⁸	-0.017	2.768×10 ⁻⁴	1	.987
	ART	-1.567	0.407	0.209	-3.846	14.793	1	<.001
	Income	1.206	0.390	3.342	3.091	9.554	1	.002

Table 4: Models' Coefficients

Model	Parameter	Estimate	Standard Error	Odds Ratio	z	Wald Test		
						Wald Statistic	df	p
M ₈	(Intercept)	8.697	1032.283	5983.125	0.008	7.098×10 ⁻⁵	1	.993
	Distance	1.754	0.415	5.775	4.229	17.882	1	<.001
	Duration	2.155	0.467	8.632	4.615	21.300	1	<.001
	Housing	3.015	0.606	20.397	4.978	24.777	1	<.001
	Hchallenge	-1.878	0.482	0.853	-3.899	15.202	1	<.001
	Education	-18.345	1032.281	1.078×10 ⁻⁸	-0.018	3.158×10 ⁻⁴	1	.986
	ART	-1.882	0.444	0.152	-4.238	17.963	1	<.001
	Income	1.849	0.475	6.351	3.896	15.175	1	<.001
	Employment	1.204	0.475	3.332	2.536	6.430	1	.011
	(Intercept)	9.024	1002.038	8297.250	0.009	8.110×10 ⁻⁵	1	.993
M ₉	Distance	1.908	0.436	6.742	4.380	19.184	1	<.001
	Duration	2.162	0.468	8.692	4.622	21.364	1	<.001
	Housing	3.035	0.612	20.805	4.961	24.614	1	<.001
	Hchallenge	-1.994	0.486	0.836	-4.099	16.800	1	<.001
	Education	-	1002.0	9.363×1	-	3.404×	1	.985

Table 4: Models' Coefficients

Model	Parameter	Estimate	Standard Error	Odds Ratio	z	Wald Test		
						Wald Statistic	df	p
	n	18.487	36	0 ⁻⁹	0.018	10 ⁻⁴		
	ART	-1.737	0.446	0.176	-3.893	15.159	1	<.001
	Income	2.033	0.489	7.640	4.156	17.271	1	<.001
	Employment	1.149	0.471	3.155	2.437	5.941	1	.015
	Water	-0.770	0.433	0.463	-1.779	3.164	1	.075
M ₁₀	(Intercept)	43.374	4593.789	6.875×10 ⁺¹⁸	0.009	8.915×10 ⁻⁵	1	.992
	Distance	1.849	0.432	6.352	4.281	18.323	1	<.001
	Duration	2.065	0.464	7.886	4.454	19.838	1	<.001
	Housing	2.965	0.605	19.403	4.898	23.987	1	<.001
	Hchallenge	-1.938	0.488	0.844	-3.971	15.768	1	<.001
	Education	-18.552	998.267	8.768×10 ⁻⁹	-0.019	3.454×10 ⁻⁴	1	.985
	ART	-1.584	0.448	0.205	-3.536	12.500	1	<.001
	Income	2.108	0.489	8.228	4.312	18.593	1	<.001
	Employment	1.071	0.470	2.918	2.281	5.201	1	.023
	Water	-	0.441	0.418	-	3.904	1	.048

Table 4: Models' Coefficients

Model	Parameter	Estimate	Standard Error	Odds Ratio	z	Wald Test		
						Wald Statistic	df	p
		0.871			1.976			
	Age	17.049	2242.006	3.941×10^{-8}	0.008	5.783×10^{-5}	1	.994

Note. High HIV/AIDS mortality Risk '1' is coded as class 1.

4.1. Performance Diagnostics

The confusion matrix is shown in Table 5. At a cut-off of 0.5, the model performs as follows: Sensitivity value of 0.964 (True Positive Rate, 96.43%); thus, the model is exceptional at identifying positive cases. Out of 224 actual "1"s, it correctly identified 216. It reports a specificity value of 0.52 (True Negative Rate, 52.00%); the model performs at an average level at identifying negative cases. It misclassified 36 out of 75 actual "0"s as positives (False Positives).

Table 5: Confusion matrix

Observed	Predicted		% Correct
	0	1	
0	39	36	52.00
1	8	216	96.43
Overall % Correct			85.28

Note. The cut-off value is set to 0.5

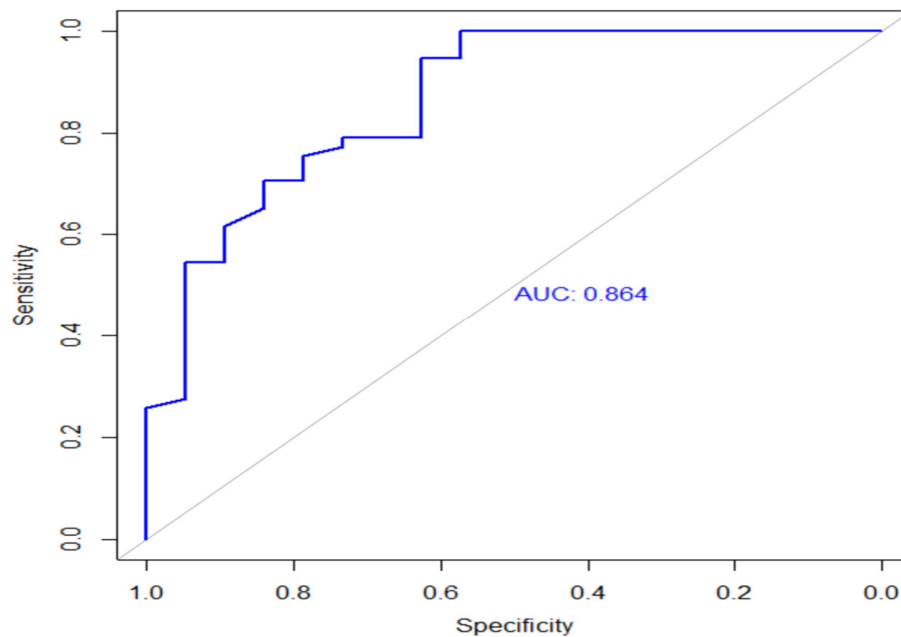
Table 6 presents the performance metrics. The model shows an AUC (Area Under the Curve) value of 0.858. This is the true measure of the model's discriminative power. An AUC of 0.858 means that if one picks a random positive and a random negative case, the model will assign a higher risk score to the positive case approximately 86% of the time. However, the model shows a specificity (52.0%) value, giving an average at identifying "No Risk" cases (Class 1), which is a limitation of the model. The F-Measure (0.908) score is high, but the H-Measure (0.474) is significantly lower, buttressing the limitation of the proposed model. The overall accuracy value of the model is 85.28%.

Table 6: Performance metrics

	Value
Accuracy	0.853
AUC	0.858
Sensitivity	0.964
Specificity	0.520
Precision	0.857
F-measure	0.908
H-measure	0.474

4.2. The ROC Curve Plot

The ROC curve plots the True Positive Rate (Sensitivity) against the False Positive Rate (1 - Specificity). The curve bows significantly towards the top-left corner, which indicates a good model. This visually confirms the high AUC value of 0.858. The area under the curve represents the overall discriminative ability between the two classes. The Area Under the ROC Curve is 0.864. This is considered "good" to "excellent" discriminative ability. It means there is an 86.4% chance that the model will rank a randomly chosen positive instance higher than a randomly chosen negative one. The model is highly effective at predicting the presence of high HIV/AIDS mortality risk (high sensitivity) but very poor at predicting its absence (low specificity).

**Figure 1: ROC plot**

4.3. Precision-Recall (PR plot) Curve

The Precision-Recall (PR plot) curve shown in Figure 2 is often more informative than the ROC curve for imbalanced datasets. The plot shows a high precision (around 0.86) maintained for a wide range of recalls (sensitivity) before it drops. The fact that the curve stays high until high recall levels indicates that the model performs exceptionally well in identifying positive instances without too many false positives, up until a point. The sharp drop at the very end suggests that to capture the final few true positives, the model starts making many more false positive errors.

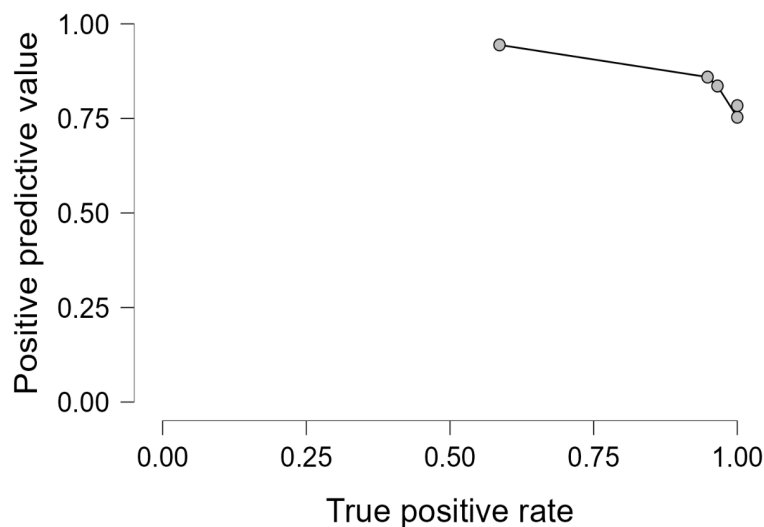


Figure 2: Precision-recall plot

4.4.1. Mixed-method Integration Design and Rationale

This study adopted a convergent parallel mixed-methods design (Creswell & Plano Clark, 2018), meaning quantitative data (structured questionnaire, $n = 300$) and qualitative data (semi-structured interviews, $n = 12$; focus group discussions, $n = 4$) were gathered at the same time, examined separately, and then brought together during interpretation. The integration served three purposes. First, corroboration—to check whether the statistical predictors of high HIV/AIDS mortality risk matched what urban residents actually experience in their daily lives. Second, insight to shed light on surprising quantitative results that the regression output alone could not account for. Third, contextualization to ground the "urban penalty" hypothesis in the specific social, economic, and infrastructural conditions of Tamale Metropolis.

4.4.2. Convergence Between Quantitative and Qualitative Findings

Five predictors showed strong agreement between the two strands:

Housing (OR = 20.40). The strongest statistical predictor came up in every focus group. Participants spoke of rooms without windows, shared toilets, and no space to isolate when someone falls ill. The qualitative data suggest housing is not just a correlate but an enabling environment, a setting that fosters both infection and poor treatment outcomes.

Duration without services (OR = 8.63). Long-term residents described a slow erosion of social support and the cumulative weight of neglect.

Income (OR = 6.35) and distance (OR = 5.78). These two predictors were described as interacting barriers. Distance only becomes prohibitive when income is low. Participants talked about the trade-off between transport costs and food, laying bare the economic calculus of seeking care. ART access (OR = 0.15). The protective effect of ART was statistically strong, but qualitative data revealed a critical caveat: access does not equal uptake. Stigma, visible queues, and fear of disclosure kept people from consistent engagement with care even when treatment was available. This "convergence-with-caveat" strengthens the reading that ART's protective effect depends on social and institutional conditions.

4.4.3. Explaining Counterintuitive Findings

Three quantitative findings were counterintuitive and needed qualitative explanation: Health challenges (OR = 0.85). The negative association—suggesting that reporting health challenges reduced mortality risk—was initially puzzling. Qualitative data pointed to a selection effect: those who report health challenges are already engaged with the health system and therefore on treatment, while those who do not report may simply be undiagnosed.

Employment (OR = 3.33). Employment appeared to increase mortality risk, contrary to expectations. Qualitative data showed that "employment" in this context is largely informal, mobile, and irregular—traders who travel, sleep away from home, and have no time for clinic visits. Employment increases exposure while reducing care access, which explains the elevated odds. Income (OR = 6.35). Higher income also appeared to increase risk, which was unexpected. Qualitative data revealed a dual effect: income enables mobility and social mixing (increasing exposure) while also enabling access to care (reducing risk). In this sample, the exposure effect appears to dominate. This finding underscores the need for disaggregated income and job-type data in future models.

4.4.4. Stigma as a Missing Mediator

Stigma did not emerge as a statistically significant predictor in the final model (OR = 1.16, $p = 0.84$). Yet it dominated qualitative accounts across all focus groups and interviews. This divergence suggests that stigma operates indirectly through delayed testing, poor adherence, and avoidance of care rather than as a direct driver of mortality risk.

Stigma is likely a mediator or effect modifier rather than an independent predictor. Future models should test stigma as a mediator between ART access and mortality risk, or as an effect modifier of the distance–mortality relationship. The qualitative strand thus identifies a variable that the quantitative model could not capture, a classic contribution of mixed-methods integration. Housing is not just a physical structure but a social space where privacy, dignity, and care are impossible. Distance is not just kilometres but a lived experience of flooded roads, unsafe travel, and lost wages. Income is not just a number but a determinant of survival where transport, food, and treatment compete. ART access is not just availability but a negotiated decision shaped by stigma, disclosure, and community visibility. Duration is not just time but accumulated exclusion and eroded social support.

The urban penalty, therefore, is not a single mechanism but a convergence of structural, economic, social, and institutional pathways that jointly amplify mortality risk. The counterintuitive findings for health challenges, employment, and income were not errors but context-dependent effects that required qualitative explanation.

5. Firth's penalized likelihood logistic Regression

Table 7 shows the output for Firth's penalized likelihood logistic regression. To resolve the separation problem (infinite estimates, massive SEs) present in the standard `glm()` output, Firth's penalized likelihood logistic regression is further used. The model is highly statistically significant. The likelihood ratio test shows a chi-square value of 104.71 at 15 degrees of freedom. The Wald test also shows a chi-square value of 74.71. These tests are all statistically significant. Firth's method successfully resolved the separation problem that made Age and Education uninterpretable in standard logistic regression. Education emerged as a strong protective factor ($OR \approx 0.012$, $p < 0.001$) — it should be retained in the final model. Age remains non-significant ($p = 0.396$) even after correction. Eight predictors are significant: Education, Employment, Income, Duration, Housing, Distance, ART access, and Hchallenge.

The model is globally significant (LR test $p < 0.001$). Counterintuitive signs persist for Income, Employment, and Hchallenge; these need further investigation.

Table 7: Coefficients of Firth's logistic regression

	coefficient	Se (coef)	Lower 0.95	Upper 0.95	Chi- square	p	Method
intercept	-1.7408	4.1464	-9.4220	9.2814	0.1532	6.955E-01	2
Age	-12277	1.5658	-6.2196	1.3318	0.7220	3.954E-01	2
Gender	-0.2492	0.4063	-1.0870	0.5659	0.3558	5.508E-01	2
Education	-4.3972	1.4799	-9.3208	-2.1131	21.7785	3.060E-06	2
Employment	1.1211	0.4541	0.2192	2.0778	5.9789	1.448E-02	2
Income	1.9796	0.4676	1.0642	2.9587	18.7407	1.497E-05	2
Dration	2.2026	0.4853	1.2643	3.2438	23.2411	1.429E-06	2
Housing	2.9596	0.5930	1.8281	4.2646	31.9654	1.569E-08	2
Water	-0.7731	0.4236	-1.6542	0.0545	3.3428	6.750E-02	2
Distance	1.7212	0.4254	0.8943	2.6097	17.3039	3.185E-05	2
Awareness	-0.4415	0.6798	-1.8660	0.8721	0.4256	5.141E-01	2
Tested	-0.4145	0.5424	-1.5413	0.6482	0.5713	4.497E-01	2
ART	-1.4696	0.4536	-2.4411	-0.5916	11.3698	7.465E-04	2
Stigma	0.1512	0.7121	-1.3261	1.5707	0.0433	8.352E-01	2
Hchallenge	-1.7973	0.4754	-2.8186	-0.8789	15.8373	6.903E-05	2

NOTE:

Method: 1-Wald, 2-profile penalized log-likelihood, 3-none

Likelihood ratio test = 104.7106 on 15 df, $p = 1.6653E-15$, $n=299$

Wald test = 74.7067 on 15 df, $p = 6.3964E-10$

6. Conclusion

This study successfully developed and evaluated a predictive logistic regression model for assessing mortality risk factors associated with urbanization and HIV/AIDS in Tamale Metropolis. The findings provide compelling evidence that urbanization-related structural determinants significantly influence HIV/AIDS outcomes, extending beyond traditional biomedical risk factors. The final model (M_8) worked well as a screening tool, detecting 96.4% of high-risk cases with an AUC of 0.864. It is built for detection, not for ruling out risk, which is acceptable when missing a case is costly. Housing (OR = 20.40) shows up as the strongest factor. Informal housing, poorly ventilated rooms, shared toilets, and lack of isolation space turn shelter into a pathway for infection and poor treatment outcomes. Duration without services (OR = 8.63), which long-term residents described as eroding social support and cumulative neglect. Income (OR = 6.35) and distance (OR = 5.78) interact since transport costs compete with food and treatment, forcing impossible choices. ART access (OR = 0.15) also indicates strong protection, but access remains a problem. Stigma, visible queues, and fear of disclosure keep people away. Findings of employment and higher income appeared to increase risk.

Furthermore, informal, mobile trading increases exposure while leaving no time for clinic visits; income enables mobility and social mixing, and the exposure effect dominated. Stigma dominated interviews but was not a direct predictor. It likely works indirectly through delayed testing, poor adherence, and care avoidance as a mediator.

The consequence is not one root cause but a convergence of structural, economic, social, and institutional pathways. Housing is a social space; distance is worsened by bad roads and low wages; income is survival whilst duration is accumulated exclusion. Structural factors are active drivers, not background context. ART's success depends on addressing barriers to consistent access and adherence.

The study contributes to the growing body of evidence suggesting that effective HIV/AIDS interventions in rapidly urbanizing African cities must address upstream determinants of health. The success of ART in reducing mortality is dependent on addressing the structural barriers that prevent consistent access and adherence. The evidence from this study underscores an urgent need to reduce HIV/AIDS mortality in Tamale Metropolis; health interventions must transcend clinical care and address the fundamental structural inequities created by rapid urbanization. The developed predictive model provides an actionable tool, but its utility ultimately depends on the political will to invest in urban health infrastructure, housing improvement, and socioeconomic development. As Tamale continues its trajectory of urban growth, integrating health considerations into urban planning is not merely desirable; it is essential for achieving Sustainable Development Goals 3 (Good Health and Well-being) and 11 (Sustainable Cities and Communities).

We recommend that the Ghana Health Service consider combining mobile clinics with stigma reduction, or pairing housing upgrades with community sensitization, which are more likely to succeed than those based on statistical associations alone.

The limitations of the study include that data were collected in parallel rather than sequentially; qualitative findings could not be used to refine the quantitative instrument.

Competing Interest

The authors do not have any conflict of interest in carrying out this study.

Funding

The authors did not receive any funding in carrying out this study

References

- [1] Abdul-Samed, F. G., Abubakari, A., Yussif, B. G., Aninanya, G. A. (2024). Determinants of adherence to antiretroviral therapy among people living with HIV receiving care in health facilities in Tamale Metropolis, Ghana. *BMC Infect Dis.* 24(1):1379.
- [2] Akinwale, A. A. (2023). Gender equity and social progress: Empowering women and girls to drive sustainable development in sub-Saharan Africa. *International Journal of Innovation Research and Advanced Studies*, 2 (2), 131- 153.
- [3] Argaw, G. S., Gelaye, K. A., Lakew, A. M., Aragaw, F. M., Chekol, Y. M., Tesfie, T. K., Gelaw, N. B. (2024). Survival and predictors of mortality among HIV-infected adults after initiation of antiretroviral therapy in Eastern Ethiopia Governmental hospitals, from January 2015 to December 2021 (multi-center retrospective follow-up study). *BMC Infect Dis.*;24(1):1352. doi: 10.1186/s12879-024-10225-2. PMID: 39592986; PMCID: PMC11600927.
- [4] Cochran, W. G. (1977). *Sampling Techniques* (3rd ed.). John Wiley & Sons, New York
- [5] Dwomoh, D., et al. (2018). Review of errors in the issue of medical certificates of cause of death in a tertiary hospital in Ghana. Published in the *International Journal of Medicine and Medical Sciences* and indexed.
- [6] Emmanuel Mensah, C., Mbiyeiba, J., & Adiyaga Wezena, C. (2022). Perception of HIV/AIDS among Tertiary Education Students in the Tamale Metropolis and Their Willingness to Take the HIV Test. *Journal of Advances in Medicine and Medical Research*, 34(7), 7–16. <https://doi.org/10.9734/jammr/2022/v34i731322>
- [7] Fuseini, I. (2016). Characterising urban growth in Tamale, Ghana: An analysis of urban governance response in infrastructure and service provision. *Habitat International*, 56, 109-123. <https://doi.org/10.1016/J.HABITATINT.2016.05.002>.
- [8] Ghana Statistical Service (GSS) (2021). *2021 Population and Housing Census General Report. Accra, Ghana: GSS.*

- [9] Kusturica, M. J. (2023). Urban environment, unhealthy buildings as a root of inequalities and importance of SDGs. *European Journal of Public Health*, 33. <https://doi.org/10.1093/eurpub/ckad160.714>.
- [10] Linganisio, W. S. and Gwegweni, J. M. T. (2016). What Perpetuates the Spread of HIV/AIDS in Rural South African Communities? A Closer Look at Social Factors. *Austin Journal of Public Health Epidemiology*. 3(1): 1031. ISSN: 2381-9014.
- [11] Marie, J. and Iqbal, B. A. (2024). Infrastructure development, informal economy, and gender inequality in Sub-Saharan Africa. *Journal of Infrastructure, Policy and Development*, 8 (6), 3413.
- [12] Marzukhi, M. G. (2023). Assessing the Effect of High-Density Housing on Mental Health of Urban Population. Case Study: Federal Territory of Kuala Lumpur. IOP Conference Series. *Earth and Environmental Science*, 1217. <https://doi.org/10.1088/1755-1315/1217/1/012020>.
- [13] Musonda E, Phiri M, Shasha L, Bwalya C, Musemangezhi S, Ishimwe SMC, Kalinda C. (2023) Prevalence of HIV testing uptake among the never-married young men (15-24) in sub-Saharan Africa: An analysis of demographic and health survey data (2015-2020). *PLoS One*. 5;18(10): e0292182. doi: 10.1371/journal.pone.0292182. PMID: 37796957; PMCID: PMC10553359
- [14] Musonda, E. P. (2023). Geographical variations and factors associated with recent HIV testing prevalence in Ghana: spatial mapping and complex survey analyses of the 2014 demographic and health surveys. *BMJ Open*, 11. <https://doi.org/10.1136/bmjopen-2020-045458>.
- [15] Nutor, J. J. (2021). Prevention of mother-to-child transmission of HIV in Ghana: Examining the gaps and risks. *Journal of Nursing Scholarship*, 53(4), 466-474.
- [16] Olawade, D. B. (2024). Exploring the current trends in social epidemiology in Sub-Saharan Africa. *Public health in Sub-Saharan Africa*, Routledge, 3-20
- [17] Sefah, I. A., Jelić, A. G., Kurdi, A., and Godman, B. (2022). Barriers and facilitators of adherence to antiretroviral treatment at a public health facility in Ghana: a mixed-method study. *Hospital Practice*, 50(2), 118-128.
- [18] Sulemana, I., Nketiah-Amponsah, E., Codjoe, S. N. A., & Andoh, J. (2019). Urbanization and income inequality in Sub-Saharan Africa. *Sustainable Cities and Society*, 48, 101544.
- [19] Ullah, S., Naz, A., Ali, B., and Ul Hakim, A. (2021). Socio-cultural factors and the transmission of HIV/AIDS in Malakand Division: a quality analysis. *Clinical Social Work and Health Intervention*, 12(2): 17-24. DOI: 10.22359/cswhi_12_2_03.

Appendix

Questionnaire On Modelling The Mortality Risk Factors Associated With Urbanization And Hiv/Aids In Tamale Metropolis

Sample Questionnaire for Urban Residents & HIV/AIDS Patients

Section 1: Demographics

1. Age: _____
2. Gender: ☐ Male ☐ Female
3. Educational Level: ☐ No formal education ☐ Primary ☐ Secondary ☐ Tertiary
4. Employment Status: ☐ Employed ☐ Unemployed ☐ Self-employed ☐ Student
5. Monthly Income Level: ☐ Low (GHc300-GHc500) ☐ Medium (GHc500-GHc1200) ☐ High (GHc1500+)
6. How long have you lived in this urban area? _____

Section 2: Urbanization & Health

7. What is your primary housing type?
☐ Permanent house ☐ Temporary structure
8. Do you have access to clean drinking water? ☐ Yes ☐ No
9. How far is the nearest healthcare facility from your residence?
☐ Less than 1 km ☐ 1-5 km ☐ More than 5 km
10. Have you or your household members experienced any major health challenges in the past year? ☐ Yes ☐ No
 - If yes, please specify: _____

Section 3: HIV/AIDS Awareness & Access

11. Are you aware of HIV/AIDS and its transmission methods?
☐ Yes ☐ No
12. Have you ever taken an HIV test? ☐ Yes ☐ No
13. If HIV-positive, are you currently receiving antiretroviral therapy (ART)? ☐ Yes ☐ No
14. If no, what are the main challenges preventing access to ART? (Check all that apply)
☐ Distance to health facility ☐ Cost of treatment ☐ Stigma ☐ Lack of awareness
15. Do you believe there is stigma associated with HIV/AIDS in your community? ☐ Yes ☐ No

Section 4: Public Health & Mortality Perception

16. In your opinion, what are the main health challenges in urban areas?
☐ HIV/AIDS ☐ Malaria ☐ Poor sanitation ☐ Lack of healthcare access ☐ Other (Specify)
17. Do you think urbanization has increased health risks? ☐ Yes ☐ No
 - If yes, how? _____
18. What improvements do you think are needed to reduce health risks in urban areas?