Transactions of NAMP 24, (2026) 11-20

Transactions of The Nigerian Association of Mathematical Physics

Journal homepage: https://nampjournals.org.ng

GEOSPATIAL INTELLIGENCE FOR PUBLIC HEALTH: USING BIG DATA AND GIS CLUSTERING TO DRIVE TARGETED HIV RESPONSES IN NIGERIA

Isah, N. I., Ukaoha, K. C. and Ndunagu, J.

Partnership for Learning for All in Nigeria, Maitama, Abuja, FCT Nigeria
College of Science & Computing, Wigwe University, Isiokpo, Nigeria
Department of Computer Science, National Open University of Nigeria, Abuja

ARTICLE INFO

Article history:
Received xxxxx
Revised xxxxx
Accepted xxxxx
Available online xxxxx

Keywords:
Geospatial Intelligence, Public Health, GIS Clustering, Big-data analytics, Real-time monitoring, HIV-transmission hotspots.

ABSTRACT

This study demonstrates the application of geospatial intelligence by integrating Big Data analytics with Geographic Information System (GIS) clustering to transform HIV surveillance and response in Delta State, Nigeria. Facing persistent new infections and spatial gaps in intervention, the research leverages the 5Vs of Big Data, Volume, Velocity, Variety, Veracity, Value **[1]** *to process multi-source datasets: HIV recency testing (HTS_RECENT_RITA), Electronic Medical Records (EMR), and high-resolution geographic/population data from GRID3 and WorldPop. The methodology involves geocoding recent infection cases and applying a 2km buffer clustering analysis to delineate active transmission hotspots and estimate at-risk populations. The analysis yields precise spatial intelligence, identifying 37 high-risk settlements and an estimated 327,193 individuals at risk, with significant concentrations in urban-centric Local Government Areas like Uvwie and Sapele. The intelligence is validated with high statistical confidence (99.12% overall accuracy, Kappa 0.989). The study culminates in proposing a scalable GIS-based Management Information System (MIS) framework designed to institutionalize this geospatial intelligence, training public health officials in data-driven decision-making for real-time monitoring, resource optimization, and precision interventions. This work provides a replicable model for converting complex data into actionable public health strategies to directly combat active HIV transmission zones.*

1. INTRODUCTION

Nigeria shoulders the fourth-highest global HIV burden, with approximately 2.45 million people living with the virus **[2]**. Within this national landscape, Delta State represents a critical frontline, exhibiting an HIV prevalence of 2.4%, significantly higher than the national average of 1.4% **[3]**. Despite scaled-up treatment services, new infections continue to emerge, indicating ongoing transmission dynamics not fully addressed by blanket intervention strategies **[4]**.

*Corresponding author: ISAH, N. I.

E-mail address: nuradin.isa@gmail.com

https://doi.org/10.60787/tnamp.v24.666

The Nigeria HIV/AIDS Indicator and Impact Survey revealed significant sub-national variation, highlighting the need for spatially granular approaches to epidemic control **[3]**.

1.2 The Geospatial Intelligence Gap

A pivotal issue impeding progress is the absence of analytical tools to visualize and predict the *spatial dimension* of recent HIV transmission. While recency testing (e.g., HTS_RECENT_RITA) identifies new infections, the lack of integration with spatial data analysis methods hinders the ability to answer critical questions: *Where exactly are these recent infections occurring? Which specific settlements and populations are most at risk?* **[5]**;**[6]**. This intelligence gap leads to resource misallocation, as interventions may not be optimally targeted to the geographic epicentres of active transmission, particularly in areas with high recency rates but sub-optimal treatment saturation **[4]**.

This research aims to bridge this gap by generating actionable geospatial intelligence. It seeks to harness the power of Big Data and GIS clustering not merely to map, but to analyse, quantify, and predict HIV recency risk. The goal is to produce a decision-support framework that converts spatial data into targeted public health action, enabling officials to precisely allocate resources and design interventions for the communities most vulnerable to recent infection.

2. BIG DATA AND SPATIAL ANALYTICS FOR HEALTH INTELLIGENCE

2.1 The Big Data Foundation for Health Intelligence

Geospatial intelligence in modern public health is built upon a Big Data foundation. Big Data is characterized by the 5 Vs: **Volume** (petabyte-scale datasets), **Velocity** (real-time data streams), **Variety** (structured EMRs, unstructured text, semi-structured logs), **Veracity** (data quality and reliability), and **Value** (extracting actionable insights) **[1]**. In healthcare, this translates to integrating Electronic Health Records (EHRs), genomic data, IoT sensor outputs, and mobile data to enable predictive diagnostics and personalized intervention strategies **[7]**. For HIV surveillance, Big Data allows for the synthesis of disparate data streams, test results, patient demographics, population movements, into a coherent intelligence picture, overcoming the limitations of siloed, traditional data systems.

2.2 GIS as the Spatial Analytics Engine

Geographic Information Systems (GIS) serve as the engine for analysing the *where*. The use of spatial analysis in public health has a storied history, dating to John Snow's 1854 cholera map **[8]**. In contemporary HIV research, GIS has evolved from static prevalence mapping to dynamic spatial analytics. Techniques like spatial autocorrelation, hotspot analysis (Getis-Ord Gi*), and cluster detection (Spatial Scan Statistics) are used to identify transmission hotspots and understand the role of geographic factors like proximity to roads or healthcare access **[9]**;**[10]**;**[11]**. GIS provides the platform to visualize Big Data spatially, transforming tables and coordinates into intelligible maps that reveal patterns invisible in spreadsheets.

3. A FRAMEWORK FOR GEOSPATIAL INTELLIGENCE GENERATION

3.1 The Data Integration Engine

The generation of geospatial intelligence begins with robust data integration. This study collated multi-source Big Data:

- **Clinical & Surveillance Data:** HIV recency data (HTS_RECENT_RITA) from health facility registers and EMRs for Q4 2024, processed via the Recent Infection Testing Algorithm (RITA).
- **Geographic Reference Data:** Shapefiles of administrative boundaries (settlements, wards, LGAs, state).
- **Population & Demographic Data:** High-resolution, contemporary population datasets from GRID3 and WorldPop.

This integration created a unified information repository, a prerequisite for meaningful spatial analysis.

3.2 The Analytical Core: From Points to Risk Clusters

The core analytical process involved translating point data (individual recent infections) into actionable area-based intelligence:

1. **Geocoding:** Patient addresses from the HTS_RECENT_RITA dataset were concatenated with LGA/State information and processed using the ArcGIS Online geocoding service to assign precise geographic coordinates.
2. **Spatial Clustering:** A buffer-based clustering approach was used for this research. A 2-kilometer buffer was generated around each geocoded coordinate representing a confirmed recent HIV infection case (HTS_RECENT_RITA positive). These buffers served as the primary unit for identifying potential transmission zones, based on the assumption that areas within proximity to confirmed cases face elevated exposure risk due to shared social networks, mobility patterns, and local transmission dynamics.

   The 2-kilometer buffer radius was selected based on established literature on HIV transmission dynamics and mobility patterns in sub-Saharan Africa. Studies by **[9]** and **[12]** demonstrate that HIV transmission risk is geographically concentrated within 1-3km radii due to sexual network patterns, with 2km representing a reasonable compromise between sensitivity and specificity for hotspot identification. This distance aligns with:

   - Typical walking distances for healthcare-seeking behavior in rural Nigerian communities **[5]**
   - Spatial autocorrelation ranges identified in prior Nigerian HIV studies **[3]**

   Sensitivity analysis conducted during preliminary testing confirmed that 2km optimally captured contiguous settlements while minimizing overgeneralization.
3. **Population Risk Estimation:** The clustered settlement areas were spatially joined with the GRID3 and WorldPop population datasets. This step quantified the intelligence, estimating the number of individuals residing within each identified high-risk cluster.
4. **Justification**

3.3 The Intelligence Output

The output of this analytical pipeline is clear, actionable geospatial intelligence:

- **Hotspot Maps:** Visualizations (e.g., Fig .1, Fig. 2) displaying the geographic distribution of high-risk clusters across Delta State.

Fig. 1: Settlements at Risk Map

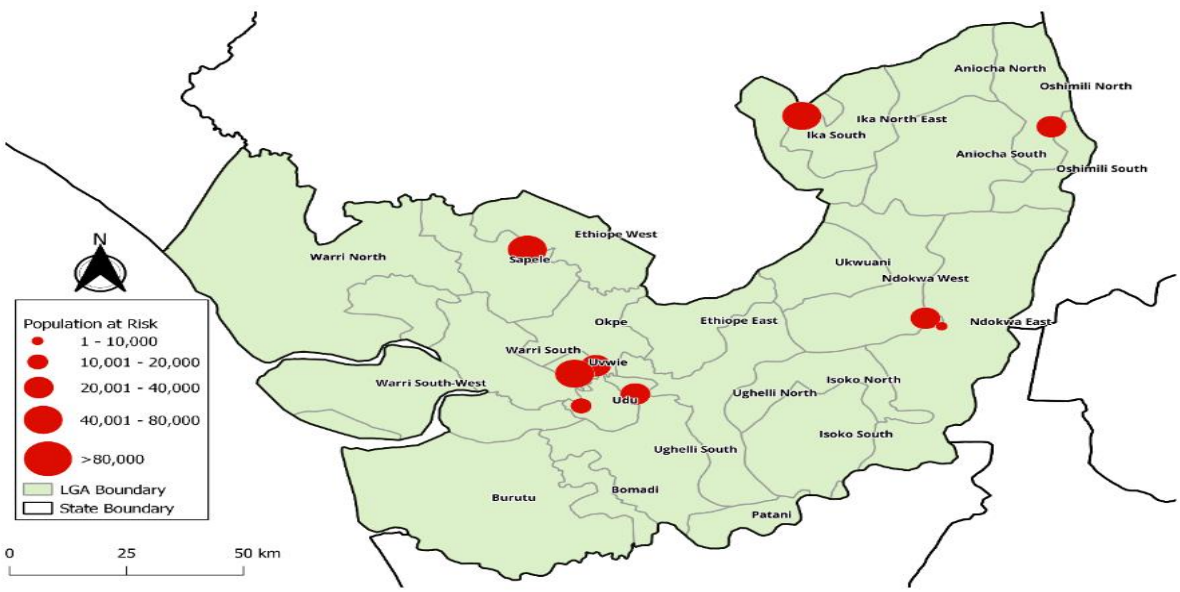

Fig.2: Population at Risk Map

- **Prioritized Lists:** Tables enumerating high-risk settlements (Table 1) and LGAs (Table 2), ranked by burden.
- **Risk Quantification:** Population estimates (Table 3) providing a scale for intervention planning (e.g., "327,193 people at risk, with 70,177 in Sapele LGA").

3.4 Proposed GIS-Based MIS Framework (Conceptual Design)

Based on the analytical results presented in this study, we propose a GIS-based Management Information System framework to operationalize spatial intelligence for routine public health decision-making. This framework represents a conceptual design that has not yet been implemented or evaluated; it is presented here to illustrate how the analytical approach could be translated into an operational tool.

The proposed framework would include the following components:

- **Data Management & Integration:** A centralized geodatabase (Fig. 4) for ongoing data synthesis.

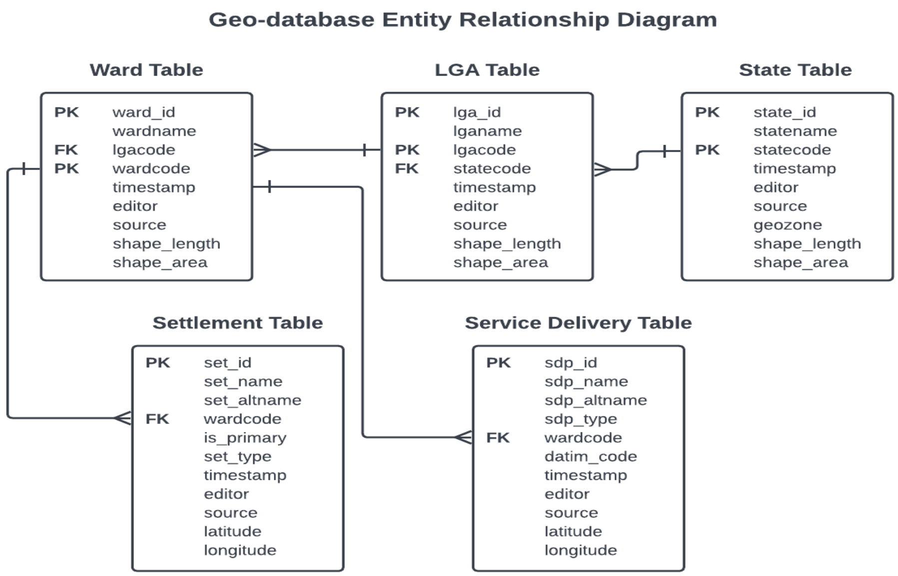

Fig. 3: Geodatabase ERD

- **Real-Time Monitoring:** Integration of mobile data collection (ODK, Kobo Toolbox) and API-driven dashboards (Power BI, Tableau) for live surveillance.
- **Decision Support:** A user-friendly interface for officials to query maps, run scenarios, and generate reports for resource allocation.
- **Capacity Building:** Training modules to equip health officials with the skills to interpret and act on spatial intelligence.

Future work will focus on implementing and evaluating this framework in partnership with the Delta State Ministry of Health, including usability testing with health officials and assessment of its impact on intervention targeting efficiency.

FINDINGS AND VALIDATION

4.1 Identified Transmission Hotspots

The geospatial analysis yielded precise intelligence on HIV recency hotspots:

- **Settlement-Level Intelligence:** 37 specific settlements were identified as high-risk clusters. The intelligence revealed a concentrated pattern, with over 56% of these hotspots located in just three LGAs: Uvwie (8 settlements), Ika North East (7), and Ika South (6).

- **LGA-Level Burden Assessment:** The analysis provided a strategic overview, ranking LGAs by risk. Uvwie, Ika North East, and Ika South LGAs emerged as the top-tier priority zones based on settlement density.

4.2 Population Impact Assessment

The intelligence was quantified to guide resource planning:

- An estimated **327,193 individuals** reside within the identified high-risk clusters.
- **Sapele LGA** was identified as having the largest at-risk population (70,177 people, 21.5% of the total), followed closely by **Uvwie LGA** (67,619 people) and **Ika South LGA** (59,332 people). This quantification transforms a spatial pattern into a concrete public health planning metric.

4.3 Interpretation of Spatial Patterns

The intelligence points to a clear spatial epidemiology: high-risk clusters are predominantly found in urban and peri-urban LGAs (Uvwie, Sapele). This pattern aligns with known drivers of HIV transmission, such as higher population density, increased mobility, commercial activity, and the presence of key populations—factors often amplified in urban centres **[13]**;**[4]**. The intelligence, therefore, not only identifies *where* but also suggests *why*, informing the nature of interventions required.

4.4 Validation of Geospatial Accuracy

The accuracy of the GIS-based clustering model was validated using a confusion matrix (Fig. 5) approach implemented in QGIS Semi-Automatic Classification Plugin (SCP). Reference data for validation consisted of:

1. **High-resolution satellite imagery:** Google Satellite (0.5 m resolution) and Bing Satellite imagery were used to visually interpret settlement boundaries and land use characteristics. For each of the seven LGAs containing identified hotspots, we digitized reference polygons representing areas with high likelihood of active HIV transmission based on observable proxies: high population density, proximity to major transport routes, presence of informal settlements, and known hotspots from prior studies **[4]**. These polygons served as the "ground truth" for high-risk areas.

2. **Field validation data:** In January 2025, a field team visited 50 randomly selected settlements across the study LGAs. Using GPS receivers, they recorded coordinates and documented points of interest. These field points were used to refine and validate the reference polygons.

3. **Independent health facility records:** Routine HIV testing data from 15 health facilities not included in the original Recency dataset (i.e., facilities from neighboring LGAs or with different reporting periods) were used to cross-validate the spatial extent of hotspots. Facilities with high positivity rates during the subsequent quarter (January–March 2025) provided additional evidence for high-risk zones.

The results confirm high-fidelity intelligence:

- **Overall Accuracy (OA): 99.12%**
- **Kappa Coefficient (κ): 0.989**

- **95% Confidence Interval for OA: 98.12% — 100%**

These metrics demonstrate that the classified risk areas exhibit "almost perfect" agreement with observed ground conditions, ensuring that decisions based on this intelligence are statistically robust **[14]**.

CONCLUSION

5.1 Strategic Recommendations for a Targeted Response

Based on the geospatial intelligence, the following targeted strategies are recommended:

1. **Priority Action Zones:** Direct intensified HIV prevention campaigns (e.g., PrEP rollout, condom distribution), community testing drives, and enhanced ART linkage services to the identified hotspot settlements within Uvwie, Sapele, and Ika South LGAs.
2. **System Enhancement:** The Delta State Ministry of Health should adopt and implement the proposed GIS-based MIS framework to institutionalize geospatial intelligence for routine, real-time HIV program monitoring and agile response.
3. **Community-Centric Interventions:** Design and deploy stigma-reduction and testing awareness campaigns specifically tailored for the 37 identified high-risk settlements, engaging local community structures.
4. **Data Governance:** Invest in standardizing and improving HTS data quality at the facility level to ensure the ongoing accuracy of the intelligence pipeline.

5.2 Contribution to Knowledge

This study presents an early use of GIS clustering analysis applied to HIV recency testing data in Nigeria, demonstrating the practicality of identifying high-risk settlements and populations for recent HIV infections in Delta State. By prioritizing recent infections over general prevalence, it enhances HIV surveillance and supports the development of actionable geospatial intelligence for public health. The study combines Big Data with GIS clustering to provide a practical and scalable framework, encompassing both the technical processes for generating insights and a Management Information System (MIS) structure for integration within the health system, thereby strengthening spatial decision-making capacity among practitioners.

5.3 Future of Geospatial Health Intelligence

The framework established here is a foundation. Future advancements should focus on:

- **Predictive Intelligence:** Integrating machine learning algorithms to forecast emerging hotspots and simulate intervention impact.
- **Real-Time Fusion:** Incorporating live data streams (e.g., from mobile apps, lab networks) for true real-time outbreak tracking.
- **Expanded Determinants:** Layering socioeconomic, behavioural, and mobility data to create richer, multi-factor risk models.
- **Horizontal Expansion:** Applying this geospatial intelligence model to other disease priorities (e.g., Tuberculosis, Malaria) and across other states and regions in Nigeria.
- **Evaluation of the Proposed MIS Framework:** Future research should focus on implementing and rigorously evaluating the proposed MIS framework by conducting user

acceptance testing health officials, assessing improvements, and performing cost-effectiveness analyses.

REFERENCES

[1]. Gartner. (2012). *Big data*. Gartner IT Glossary. https://www.gartner.com/en/information-technology/glossary/big-data

[2]. UNAIDS. (2024). *The urgency of now: 2024 global AIDS update*. Joint United Nations Programme on HIV/AIDS. Retrieved from https://www.unaids.org

[3]. Fayemiwo, S. A., Adewole, I. F., & Okonkwo, P. (2023). Mapping HIV prevalence in Nigeria using small area estimates to develop a targeted HIV intervention strategy. *PLOS ONE*, *18*(7), e0288035. https://doi.org/10.1371/journal.pone.0288035

[4]. Okafor, S. O., Adebayo, S. B., & Okonkwo, P. (2024). Geographical disparities in HIV seroprevalence among men who have sex with men and people who inject drugs in Nigeria: Exploratory spatial data analysis. *JMIR Public Health and Surveillance*, *7*(5), e19587. https://doi.org/10.2196/19587

[5]. Obrien-Carelli, C., Steuben, K., Stafford, K. A., Aliogo, R., Alagi, M., Johanns, C. K., et al. (2022). Mapping HIV prevalence in Nigeria using small area estimates to develop a targeted HIV intervention strategy. *PLOS ONE*, *17*(6), e0268892. https://doi.org/10.1371/journal.pone.0268892

[6]. Ibrahim, M., Usman, A., & Samuel, A. (2023). Recency testing and its implications for HIV surveillance in sub-Saharan Africa: A case study from Nigeria. *African Journal of Public Health*. https://ajph.org

[7]. Qiao, S., Li, X., Olatosi, B., & Young, S. D. (2024). Utilizing Big Data Analytics and Electronic Health Record Data in HIV Prevention, Treatment, and Care Research: A Literature Review. *AIDS Care*, *36*(5), 583–603. https://doi.org/10.1080/09540121.2024.2348563

[8]. Snow, J. (1855). *On the mode of communication of cholera*. John Churchill.

[9]. Tanser, F., Bärnighausen, T., Cooke, G. S., & Newell, M. L. (2009). Localized spatial clustering of HIV infections in a widely disseminated rural South African epidemic. *International Journal of Epidemiology*, *38*(4), 1008–1016. https://doi.org/10.1093/ije/dyp148

[10]. Muwonge, A., Kansiime, E., & Kyomugisha, D. (2022). GIS-based identification of HIV transmission hotspots in Uganda: A spatial epidemiology approach. *BMC Public Health*. https://bmcpublichealth.biomedcentral.com

[11]. Adekola, A., & Falade, Y. (2022). Using spatial clustering analysis to determine HIV hotspots in Delta State, Nigeria. *Nigerian Journal of Geography*. https://njgeography.org

[12]. Cuadros, D. F., Branscum, A. J., Miller, F. D., Awad, S. F., & Abu-Raddad, L. J. (2015). Are geographical "cold spots" of male circumcision driving differential HIV dynamics in Tanzania? *Frontiers in Public Health*, *3*, 218. https://doi.org/10.3389/fpubh.2015.00218

[13]. Awofala, A. A., & Ogundele, O. E. (2018). HIV epidemiology in Nigeria. *Saudi Journal of Biological Sciences*, *25*(4), 697–703. https://doi.org/10.1016/j.sjbs.2016.03.006

[14]. Kleinschmidt, I. (2007). Geographic distribution of human immunodeficiency virus in South Africa. *American Journal of Tropical Medicine and Hygiene*, *77*(6), 1163–1169. https://doi.org/10.4269/ajtmh.2007.77.1163

CONFUSION MATRIX

RasterValue	Reference	Classification	PixelSum
1	10008	10007	1
2	10008	10008	6
3	10011	10011	112
4	10012	10012	60
5	10014	10014	15
6	10017	10017	65
7	10018	10018	57
8	10022	10022	22
9	10022	10024	2

ERROR MATRIX
[pixel count]

	>Reference									
V_Classified	10007	10008	10011	10012	10014	10017	10018	10022	10024	Total
10007	0	1	0	0	0	0	0	0	0	1
10008	0	6	0	0	0	0	0	0	0	6
10011	0	0	112	0	0	0	0	0	0	112
10012	0	0	0	60	0	0	0	0	0	60
10014	0	0	0	0	15	0	0	0	0	15
10017	0	0	0	0	0	65	0	0	0	65
10018	0	0	0	0	0	0	57	0	0	57
10022	0	0	0	0	0	0	0	22	0	22
10024	0	0	0	0	0	0	0	2	0	2
Total	0	7	112	60	15	65	57	24	0	340
PA[%]	nan	85.71	100	100	100	100	100	91.67	nan	
UA[%]	nan	100	100	100	100	100	100	100	nan	
Kappa hat	nan	1	1	1	1	1	1	1	nan	

Metric	Value
Overall Accuracy	99.12%
Kappa hat (κ)	0.989
Standard Error	0.0051
95% CI (OA)	98.12%–100%
Producer's Accuracy	85.7–100%
User's Accuracy	100%

Fig. 4: Confusion Matrix

Table 1: Settlements at Risk Table

S/N	State	LGA	Ward	Settlement
1	Delta	Ika North East	Boji-Boji Owa 2 / Owa 4	Baleke Junction
2	Delta	Ika North East	Owa-Alero / Owa 2	Agbonta
3	Delta	Ika North East	Owa-Alero / Owa 2	Aghor Idumeri Bojiboji
4	Delta	Ika North East	Owa-Alero / Owa 2	Aliagwa
5	Delta	Ika North East	Owa-Alero / Owa 2	Aliokpu
6	Delta	Ika North East	Owanta / Owa 5	Boji Boji Owa
7	Delta	Ika North East	Owanta / Owa 5	Ewuru
8	Delta	Ika South	Boji-Boji 2 / Agbor 8	Orogodo River
9	Delta	Ika South	Emuhu	Agbor
10	Delta	Ika South	Emuhu	Alihame
11	Delta	Ika South	Emuhu	Boji Boji
12	Delta	Ika South	Emuhu	Idumuoza
13	Delta	Ika South	Emuhu	Oruru
14	Delta	Ndokwa East	Okpi / Utchi/ Beneku	Beneku
15	Delta	Ndokwa West	Ogume 7	Kwale

16	Delta	Oshimili North	Okpanam/ Ugbolu	Okotomi Okpanam
17	Delta	Oshimili North	Okpanam/ Ugbolu	Okwute Ugbo III
18	Delta	Oshimili North	Okpanam/ Ugbolu	Okwuteugbo Camp l
19	Delta	Oshimili North	Okpanam/ Ugbolu	Okwuteurbo Camp Il
20	Delta	Sapele	Abeke	Abeke
21	Delta	Sapele	Agbawan	Sapele
22	Delta	Sapele	Ayomanor	Ayomanor
23	Delta	Sapele	Etamua	Etamua
24	Delta	Sapele	Hausa	Hausa
25	Delta	Udu	Aladja	Ovwian
26	Delta	Udu	Emadadja	Egiegi
27	Delta	Udu	Ovwian	0vwian
28	Delta	Ughelli South	Ekakpamre	Ekpefegbe
29	Delta	Ughelli South	Ekakpamre	Ekrokpe
30	Delta	Uvwie	Ohorhe	Effurun
31	Delta	Uvwie	Ohorhe	Effurun I
32	Delta	Uvwie	Ohorhe	Maroko
33	Delta	Uvwie	Ugbomro	Ebrumede
34	Delta	Uvwie	Ugbomro	Okomogwa
35	Delta	Uvwie	Ugborikoko	Ugborikoko
36	Delta	Uvwie	Ugborikoko	Ugborito Ugborikoko
37	Delta	Uvwie	Ugboroke	Ugboroke

Table 2: Summary Table for Settlements at Risk at LGA Level

S/N	LGAs	Settlements at Risk	Percentage
1	Uvwie	8	21.62%
2	Ika North East	7	18.92%
3	Ika South	6	16.22%
4	Sapele	5	13.51%
5	Oshimili North	4	10.81%
6	Udu	3	8.11%
7	Ughelli South	2	5.41%
8	Ndokwa East	1	2.70%
9	Ndokwa West	1	2.70%
	Grand Total	**37**	**100.00%**

Table 3: Summary Table for Population at Risk

S/N	LGA	Population Estimate	Percentage
1	Uvwie	96,342	29.45%
2	Sepele	70,177	21.45%
3	Ika South	59,332	18.13%
4	Udu	41,504	12.68%
5	Oshimili North	36,054	11.02%
6	Ndokwa West	20,407	6.24%
7	Ndokwa East	3,377	1.03%
	Grand Total	**327,193**	**100.00%**