*Journal of Survey Statistics and Methodology (2026)* ***00**, 1–25*

# DEVELOPING A CUSTOMIZED, ENUMERATION AREA-BASED SAMPLING FRAME TAILORED TO A SPECIFIC POPULATION SUBGROUP USING GEOSPATIAL METHODS

SARCHIL QADER*
EDITH DARIN
AHMADOU HAMADY DICKO
HISHAM GALAL
HYUNJU PARK
REBECA MORENO JIMENEZ
ANDREW HARFOOT
ANDREW J. TATEM

A national sampling frame typically comprises a list of Primary Sampling Units (PSUs), such as enumeration areas derived from census data, which are commonly used in household surveys. Both national

SARCHIL QADER is with WorldPop Research Group, University of Southampton, Southampton SO17 1BJ, United Kingdom, and a visiting researcher with Natural Resources Department, College of Agricultural Engineering Sciences, University of Sulaymaniyah, Sulaymaniyah 46001, Kurdistan Region, Iraq. EDITH DARIN is with Leverhulme Centre for Demographic Science, University of Oxford, OxfordOX1 2JD, United Kingdom. AHMADOU HAMADY DICKO is with United Nations High Commissioner for Refugees, United Nations, Villa 7B BP, 16917 Dakar, Point E, Senegal. HISHAM GALAL is with United Nations High Commissioner for Refugees, C. Salteman 126, Panama City, Panama. HYUNJU PARK is with Global Data Service, United Nations High Commissioner for Refugees, Kastellet 2, 2100 Copenhagen, Denmark. REBECA MORENO JIMENEZ is with United Nations High Commissioner for Refugees, Rue de Montbrillant 94, 1201 Genève, Switzerland. ANDREW HARFOOT is with GeoData, University of Southampton, Southampton SO17 1BJ, United Kingdom. ANDREW J. TATEM is with WorldPop Research Group, University of Southampton, Southampton SO17 1BJ, United Kingdom.
We extend our sincere appreciation to OpenStreetMap for their invaluable efforts in digitizing significant natural and human-made features across the globe. We are also grateful to the reviewers and the associate editor for their constructive feedback and insightful suggestions, which have contributed to meaningful improvements in our original submission.
This work secured funding from the endorsed team of the Data Innovation Fund 2022–2023, which was a collaborative initiative between UNHCR's Innovation Service, Global Data Service (GDS), and Division of Information and Telecommunications (DIST).
The authors declare that they have no competing interests.
*Address correspondence to Sarchil Qader, WorldPop Research Group, School of Geography and Environmental Science, University of Southampton, University Road, Southampton SO17 1BJ, United Kingdom; E-mail: S.Qader@soton.ac.uk.

https://doi.org/10.1093/jssam/smaf027

statistical offices and non-governmental organizations often rely on this framework when conducting surveys related to forced displacement. However, these frames are generally developed without considering the estimated number or geographic distribution of displaced populations. As a result, achieving the desired sample size becomes difficult and cost-intensive, as selected units frequently contain no individuals of interest. This study aimed to evaluate the potential of geospatial methodologies to develop a digital national sample frame tailored to a specific population subgroup or the general population, with the goal of ensuring applicability across diverse settings. For the first time, this work produced publicly accessible, digitized boundaries for urban and rural areas in Cameroon that are aligned with official administrative divisions and do not follow a grid-based system. According to our classification and estimated number from the ProGres database, 46 percent of refugees in Cameroon resided in rural areas, while 31 percent lived in camps and 23 percent in urban settings. The proposed geospatial approach offers a cost-effective alternative to traditional manual methods, particularly in data-scarce environments, and eliminates common geometric inconsistencies found in manual mapping efforts. All sampling units were nested within administrative boundaries, and in populated areas, their delineations aligned with observable ground features and respected major physical barriers. Importantly, including the refugee population in the customized national sampling frame was essential, as it enhanced the representativeness of refugees within it. This approach can be easily adapted to other countries. Notably, it was implemented in preparation for 2024's Forced Displacement Survey in Cameroon, highlighting its practical application and relevance in real-world survey contexts.

## Statement of Significance

Many countries are preparing for censuses and various socioeconomic surveys. Ensuring the availability of an accurate and up-to-date national sampling frame is a critical early step in the preparation process. However, in many countries, national sampling frames are often outdated, incomplete, or unavailable. National statistical offices and international organizations allocate significant time and resources annually to develop these frames, a process that can take years and requires substantial financial investment, particularly when relying solely on conventional methods. Despite these significant efforts, most of these frameworks fail to accurately represent hard-to-reach populations, as their estimates and geographic locations are often not considered during the development of the sampling frame. This work successfully automates the development of digitized pre-enumeration areas, leveraging advanced region merging algorithms and geospatial techniques and data. The approach is both cost-effective and scalable, offering potential for global implementation.

KEYWORDS: Cameroon; Customized sampling frame; Enumeration area; Geospatial methods; Pre-EA tool; Refugee.

## 1. INTRODUCTION

Surveys and censuses are the primary sources of demographic, health, and socioeconomic data across countries. Specifically, household surveys serve as the main method for delivering timely, comprehensive health and socioeconomic data, as conducting a full census is both labor- and financially intensive. Accurate and up-to-date sampling frames are essential for population-representative household survey methodologies. A sampling frame refers to the list from which units are selected for the sample (Valliant et al. 2018). National housing and population censuses are often the primary sources of national sampling frames (Mather and Cortes 2016). However, in most countries, national housing and population censuses are typically conducted every ten years (Department of Economic and Social Affairs Statistics Division—UNSD 2017; Office for National Statistics—ONS 2022), which results in national sampling frames and census data becoming outdated. Additionally, some countries may lack digital census national sampling frames or may have incomplete ones (Blankespoor et al., 2018; Eckman and Himelein 2020), such as Somalia, Armenia, the Democratic Republic of Congo, and Libya.

The creation and updating of digital national sampling frames are a laborious process (Qader et al. 2020, 2021), primarily because many countries still rely on manual methods to digitize these frames, a process that can take years and significant financial resources to complete (Wagenaar et al. 2018). Furthermore, when constructing the census national sampling frame, only basic parameters such as overall population size, geographic areas, and local boundaries are often considered. As a result, these frames frequently omit important population groups, leading to underrepresentation when used to identify subgroups within the target population (Reichel and Morales 2017). For instance, in Colombia, Richard et al. (2011) highlighted the limitations of using census data to identify target populations, noting its unreliability in locating small, mobile populations such as illegal migrants and refugees. Therefore, it is crucial for national statistical offices and other relevant agencies to have access to the necessary geospatial techniques, tools, and datasets to enable the creation of customized national sampling frames in a timely and resource-efficient manner.

When a digital national sampling frame is unavailable or inaccessible, researchers have developed alternative methods to construct sampling frames (Gallego et al. 2015; Eckman and Himelein 2020; Miller et al. 2020; Parsaeian et al. 2021). For example, in Mozambique, open-source satellite imagery was used to create a representative community health survey sampling frame

(Wagenaar et al. 2018). Similarly, Himelein et al. (2014) demonstrated that a random geographic cluster sample (RGCS) can serve as a viable alternative to conventional methods for surveying mobile populations, such as those in Ethiopia's Afar region. However, many of these approaches are highly context-specific and have not been transformed into standalone, user-friendly tools, making them difficult to replicate.

Over the last decade, gridded population-based sampling frames have gained popularity in socioeconomic surveys for their time- and cost-saving efficiency (Galway et al. 2012; Boo et al. 2020; Thomson et al. 2020). For instance, Qader et al. (2020) used a Quadtree algorithm to develop a gridded population sampling frame for the second wave of the Somali High-Frequency Survey (Pape and Wollburg 2019). Tools such as GridSample and the Geo-sampling Tool have also been used by various organizations for different types of surveys (Thomson et al. 2017; Cajka et al. 2018). Despite their advantages, gridded sampling techniques often produce outputs that pose practical challenges in the field. Specifically, these grids may not align with recognizable ground features, leading to difficulties for enumerators and complications in calculating accurate sampling weights, as buildings and other structures may be split across grid boundaries. To address these limitations, innovative approaches that leverage the growing availability of geospatial datasets are essential for developing national sampling frames that align more closely with real-world boundaries and are practical for field implementation.

Despite the existing challenges in creating accurate national sampling frames for socio-economic surveys, developing a customized national sampling frame for vulnerable and specific population subgroups, such as forcibly displaced persons, adds additional complexities to the process. There is a growing demand for more reliable statistics on the living conditions of forcibly displaced populations, as their numbers continue to reach record levels each year (UNHCR, 2023). With the launch of the Forced Displacement Surveys (FDS) (UNHCR, 2024a) and the increasing interest in integrating forcibly displaced persons into national surveys, United Nations High Commissioner for Refugees (UNHCR) operations face a significant challenge: establishing appropriate sampling frames to draw representative samples of these populations. When available, registration data is typically used as a sampling frame for refugees and asylum seekers. However, these datasets are often based on outdated information. For internally displaced persons (IDPs), registration data is rarely available (Eckman and Himelein 2022). The standard two-stage sampling method—where small, compact geographic areas are selected at random for enumeration, followed by the selection of eligible individuals within those areas—is frequently considered too complex and costly for most operations (Miller et al. 2020). This is especially true in out-of-camp settings, where forcibly displaced populations tend to be highly dispersed and unregistered (Willis et al. 2014; Martin-Shields et al. 2019). This challenge is further exacerbated by the lack of readily available georeferenced data on the numbers and locations of displaced persons, which is critical for constructing meaningful enumeration areas. Even when such areas exist, the small proportion of displaced individuals within the general population means that most selected areas may contain no displaced persons at all—making it financially and logistically inefficient to reach the required sample size. In response to these issues, Eckman and Himelein (2022) reviewed nine different sampling strategies for interviewing refugees and IDPs. Their findings concluded that the most appropriate sampling approach depends heavily on several factors: the characteristics of the displaced population, the available data collection budget, and the researcher's tolerance for under- or over-coverage bias.

In Cameroon, there is currently no reliable and up-to-date digital national sampling frame. As of March 2023, Cameroon was home to over 480,000 refugees and asylum seekers, including approximately 349,000 from the Central African Republic and 128,000 from Nigeria, in addition to more than one million IDPs and nearly 646,000 returnees (OCHA, 2023). Conflict, violence, floods, and other sudden-onset disasters are the primary causes of displacement in the country. The country is facing a critical challenge due to the absence of an appropriate national sampling frame, which hinders the effective inclusion and representativeness of forcibly displaced populations within primary sampling units (PSUs). Given the current scale of forced displacement, failure to adequately address the needs of displaced populations will hinder progress toward the Sustainable Development Goals (SDGs) and negatively impact the peace and security of nations (IPI, 2018). Conversely, integrating refugees into SDG implementation policies can reduce the frequency and impact of displacement, while ensuring that IDPs contribute to local economies and overall economic development (IPI, 2018). Therefore, to ensure the maximum inclusion of forcibly displaced persons in future surveys, the development of a customized national sampling frame is essential.

In collaboration with various governmental and non-governmental organizations, WorldPop (https://www.worldpop.org/) and GeoData at the University of Southampton/UK developed a robust and adaptable pre-EA toolset designed to support the automated creation of enumeration area-based sampling frames (Qader et al. 2021, 2022, 2023). As a proof of concept, this project explores the tool's capacity to leverage existing geospatial datasets and its flexibility in generating a customized enumeration area-based sampling frame, specifically for surveying forcibly displaced populations (refugees) in Cameroon. The initial phase of the project focused on mapping refugee populations at a resolution of 100 meters by 100 m, utilizing modeling techniques and spatial data processing (Darin et al. 2024). These gridded refugee estimates served as a key input for the tool in determining the size and boundaries of sampling units. It is important to note, however, that this work does not evaluate the accuracy of the refugee population estimates or their geographic distribution.

## 2. MATERIAL AND METHOD

The method was designed to develop a customized enumeration area-based frame that serves as a national sampling frame specifically tailored for surveying the refugee population in Cameroon. The term “pre-enumeration” was used intentionally, as this product is not derived from census data.

### 2.1 Method Strategy

The method utilized a range of geospatial techniques and datasets to construct customized enumeration areas to be used as a sampling frame for surveying refugees in Cameroon. In our previous work, we applied statistical modelling methods in combination with informative geospatial covariates to map the refugee population, as recorded in the UNHCR-monitored registration system, at a $100 \times 100\,\text{m}$ resolution (Darin et al. 2024). These gridded refugee estimates served as the primary source for estimating both the number and spatial distribution of refugees.

In addition to population estimates, digitized images of several natural and man-made features—including roads, waterways, and administrative boundaries—were acquired, pre-processed, and harmonized using a suite of geospatial processing techniques. These datasets were then integrated into the pre-EA tool, which automated the construction of the customized enumeration area-based sampling frame.

The tool first subdivided the area of interest into discrete, easily identifiable units referred to as building blocks. For each block, it calculated and summarized key attributes such as the estimated number of refugees, their unique IDs, and the geographic area. The tool was then configured with several user-defined parameters, including population and area thresholds, uncrossable boundaries, and weighting factors, to ensure that the resulting sampling units met operational and analytical requirements. The pre-EA outputs were designed to align with visible ground features, remained nested within administrative boundaries, and respected uncrossable natural or political features. The corresponding section of this report will provide a detailed description of each procedural step.

### 2.2 Input Data Requirements

#### *2.2.1 Modelled gridded refugee population.*

Population is a key determinant—alongside other parameters—in defining the unit sizes within a national sampling frame. However, when developing a customized enumeration area-based sampling frame for refugees, more granular data is required, particularly concerning the estimated number of refugees and their geographic distribution. In the initial phase of this project, we used routinely collected refugee registration data alongside high-resolution population maps, satellite-derived settlement maps, and other spatial variables to break

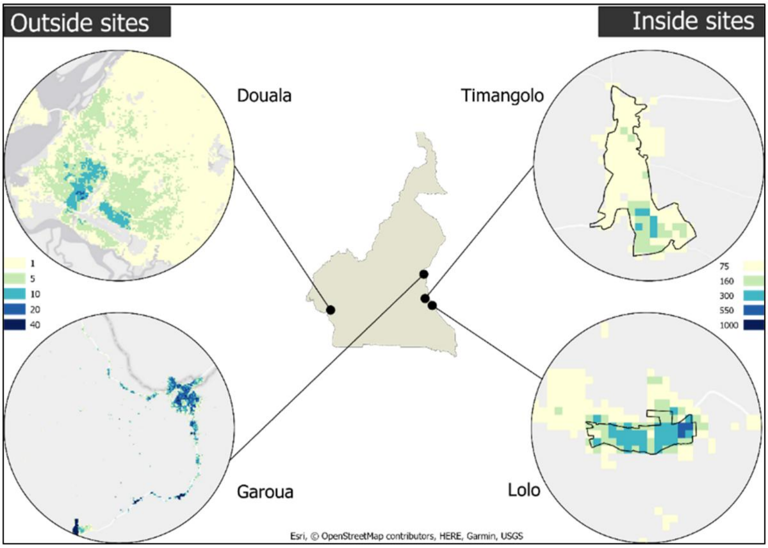


**Figure 1. Estimated gridded refugee population in Cameroon at 100 × 100 m resolution, with close-up views of selected regions to illustrate spatial distribution patterns (Darin et al. 2024).**

down total refugee counts into 100-m grid cells (Darin et al. 2024) (figure 1). Within monitored refugee sites, we assigned refugees to grid cells using a deterministic method based on the number of buildings. Outside of these sites, we applied a random forest model that uses textual geographic details from the refugee registry and high-resolution population data, as well as other geospatial covariates, to allocate refugees spatially. We demonstrated this approach using registration data from Cameroon, maintained by the UNHCR. This approach facilitated a detailed understanding of refugee distribution across the study area, which informed the geospatial tool used to develop a representative sampling frame for surveying refugees.

The full methodology for generating the gridded refugee population estimates is detailed in Darin et al. (2024).

*2.2.2 Digitised natural and man-made features.*
Traditionally, sampling unit boundaries have been manually delineated using high-resolution satellite imagery (Qader et al. 2021). During this manual process, significant effort has been invested in aligning sampling unit outlines with visible ground features to facilitate navigation and identification during fieldwork. In contrast, the automated construction process using the pre-EA tool relies heavily on existing georeferenced datasets, particularly digitized representations of natural and man-made features. For this project, digitized

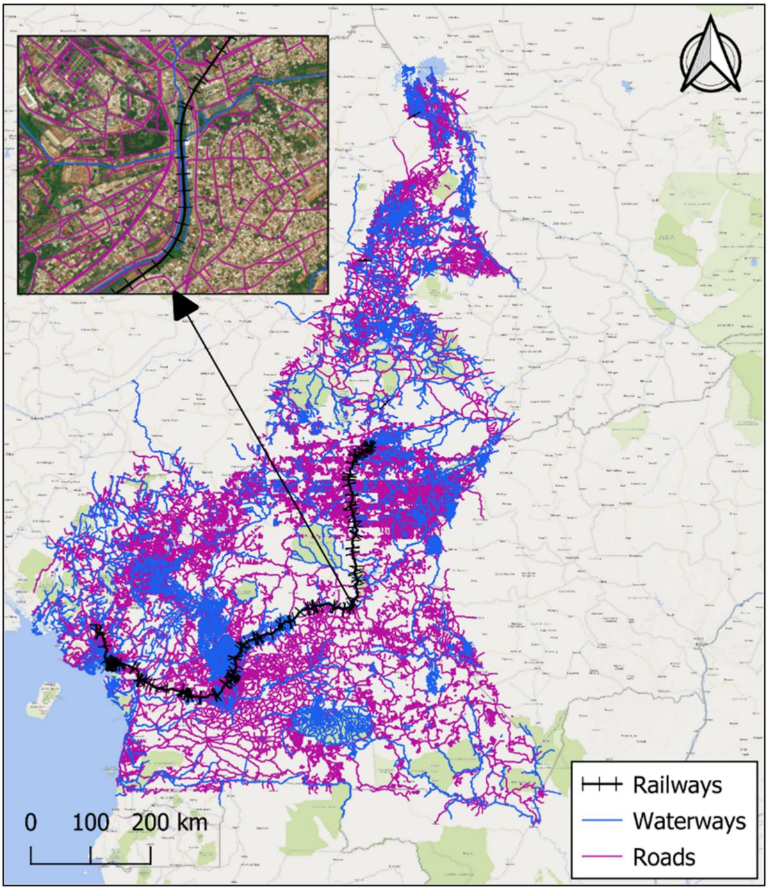


**Figure 2. Digitized roads, railways, and waterways in Cameroon, sourced from OpenStreetMap (OSM, 2023).** Basemaps displayed in QGIS are derived from Bing Maps and ESRI high-resolution satellite imagery.

boundaries such as roads, railways, and waterways were sourced from OpenStreetMap (OSM) (figure 2) (OSM 2023). OSM is a free, publicly accessible geographic database that is continuously updated and maintained by a global community of volunteers. Contributors gather data through field surveys, trace features from aerial imagery, and integrate geodata from other openly licensed sources.

*2.2.3 Administrative boundary.*

The administrative boundary is a crucial component in the creation of a sampling frame, as the outlines of sampling units must be nested within the specified administrative boundaries. For this study, the subnational administrative boundaries at the community level (ADM3) were obtained from the Humanitarian Data Exchange (HDX) (Runfola et al. 2020).

## 2.3 Data Preparation

The development of a national sampling frame is a complex process that requires careful planning and consideration (Qader et al. 2021). This process can be streamlined and potentially enhanced through the efficiency of the pre-EA tool. While the tool's application is fully automated, several preprocessing and harmonization steps were incorporated to address issues related to the coverage, quality, and consistency of the input datasets.

*2.3.1 Processing the digitised line features..*
The following preprocessing steps were applied to digitized natural and man-made features, such as roads, railways, and waterways, in preparation for the application of the pre-EA tool:

(1) All linear features, including roads, railways, and waterways, were clipped to the boundaries of Cameroon's subnational administrative level 3 (ADM3).
(2) The linear features were reprojected from World Geodetic System (WGS) 1984 to WGS 1984 UTM Zone 33 North. Projecting from WGS 1984 to a Universal Transverse Mercator (UTM) zone is necessary because UTM minimizes distortion for local-scale spatial analysis, providing more accurate measurements and better compatibility with region-specific geospatial data.
(3) Dual lines, such as those representing multiple lanes (e.g. double or triple highways), were merged into a single centerline within the road dataset to prevent the creation of complex boundary shapes in the sampling frame (see figure 3).

*2.3.2 Uncrossable features..*
When creating a digitized national sampling frame, certain natural and man-made features should be designated as uncrossable, as enumerators may be unable to traverse them during data collection. In this study, we introduced uncrossable features, and the details of these features are discussed below:

(1) **Major Roads and Rivers:** In the provided OSM road and waterway datasets, a dedicated column in the attribute table under the name "fclass" contains the types of roads and waterways. For this study, road types such as "primary" and "trunk," as well as waterway types like "river," were selected as uncrossable features (figure 4) since their traversability on the ground may be difficult or impractical.

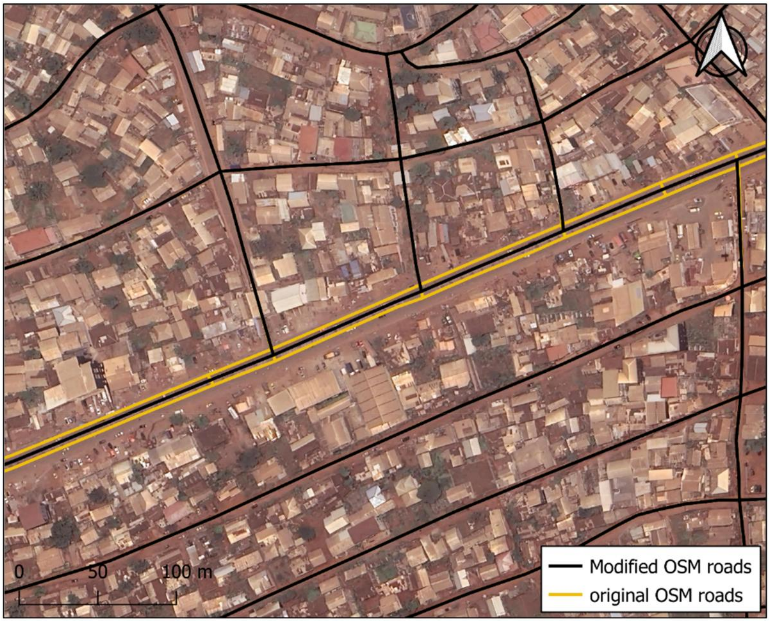


**Figure 3. Illustrates an example of parallel roads that have been merged.** The road dataset was sourced from OpenStreetMap (OSM) (2023), while the basemap was derived from ESRI satellite imagery in QGIS.

(2) **Settlement Boundaries:** To avoid mixing with neighboring areas, it is preferable for surveyors to designate the sampling units of each contiguous settled area as uncrossable within its spatial extent. This can be achieved by marking the boundaries of specific settlement areas, based on their population size, as uncrossable.
   (a) **Building Footprints:** The Digitize Africa project, conducted by Ecopia.AI and Maxar Technologies (2020), generated building footprints for 51 countries across sub-Saharan Africa. Dooley et al. (2020) calculated various building attributes and summarized them into 100m-by-100m grid cells. The locations of polygons in the building footprint dataset define the "settled" grid cells. In this study, the Gridded Maps of Building Patterns (Dooley et al. 2020) for Cameroon were polygonized to delineate the boundaries of settlements.
   (b) **Population Data:** Zonal statistics were applied to calculate the total population within each settlement boundary, based on gridded population data for Cameroon sourced from WorldPop (Bondarenko et al. 2020). An observation-based criterion was used to categorize settlements into two groups: (i) those with 500 or more people, and (ii) those with fewer than 500 people. This classification is based on the

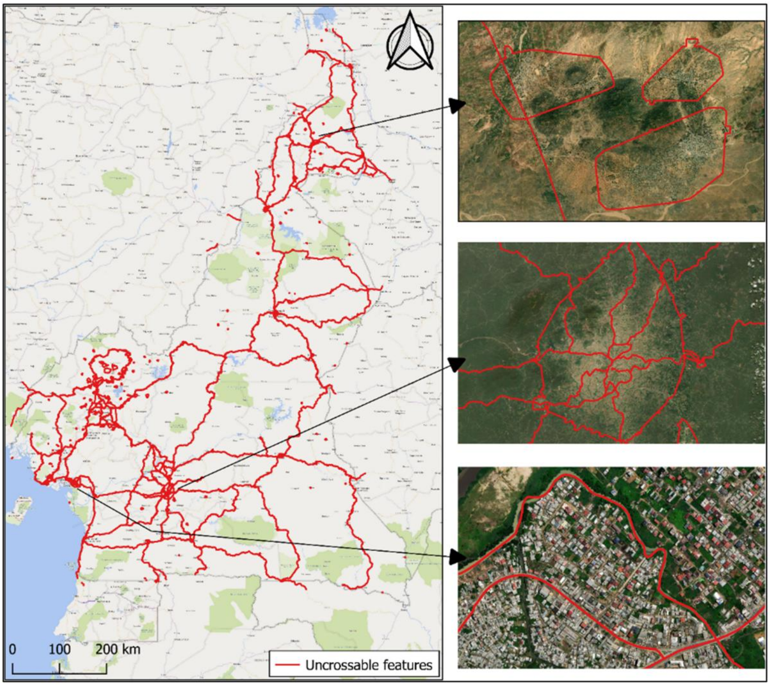


**Figure 4. Illustrates the uncrossable features generated within the customized sampling frame.** The road dataset was sourced from OpenStreetMap (OSM) (2023), while Bing Maps and ESRI high-resolution satellite imagery served as the basemap sources in QGIS.

premise that settlements with populations of 500 or more are considered major settled areas, and their sample units should not be combined with those of neighboring settlements. In the absence of local boundaries, this approach ensures that sample units from cities, towns, and larger villages remain within their spatial extents.

(3) The final uncrossable features for the country were generated by combining the selected road and waterway types with the boundaries of settlements having populations greater than 500.

*2.3.3 Input data preparation in rural areas.*

The geographic coverage of digitized natural and man-made features is often limited in rural areas. As a result, the automatic creation of sampling units may face challenges in these regions due to inconsistencies and gaps in the spatial coverage of existing digitized boundaries. The following steps outline

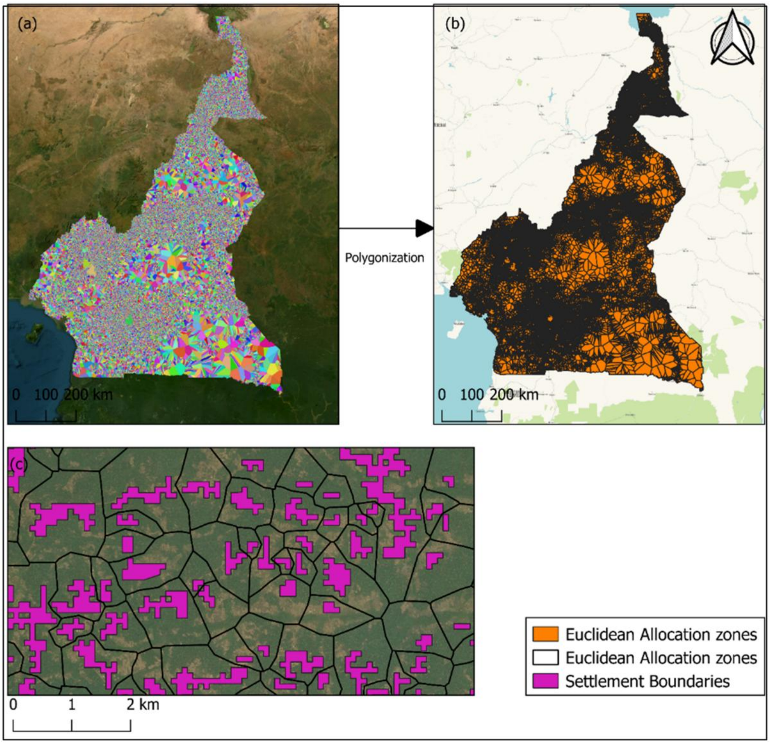


**Figure 5. Results of the Euclidean Allocation technique: (a) Raster output from the Euclidean Allocation technique (the legend for this figure is omitted due to the extensive and irrelevant list of unique values); (b) Polygonized Euclidean Allocation zone boundaries; and (c) Outline of the Euclidean Allocation boundaries surrounding settled areas, overlaid on a high-resolution satellite imagery basemap.**

the methods and techniques employed to partition rural areas in the absence of adequate digitized boundaries.

(1) **Partitioning Rural Areas in the Absence of Boundaries:** To partition rural areas where boundaries are unavailable or limited, we created envelopes (adjacent polygons) around the settled areas defined in Section 2.2.3, using the Euclidean Allocation procedure (figure 5a) (ESRI, 2025). This method assigns each grid cell to the nearest settlement boundary based on Euclidean distance. Although the primary purpose of this technique is to compute the closest source (settled area) for each cell, this study does not emphasize the distance calculation. The technique was applied across Cameroon to establish proximate zones around each settled area. Key advantages of this approach include the creation of adjacent polygons and the preservation of settlement boundaries used as inputs.

(2) **Conversion and Cropping:** The output of the Euclidean allocation process was in raster format. This raster was subsequently converted into polygons and cropped to align with the national boundary (figure 5b).
(3) **Refinement of Output:** While the Euclidean allocation technique produced adjacent polygons for all of Cameroon, it was primarily necessary for settled areas with populations smaller than 500. Therefore, the polygons representing settlements with populations over 500 were used to exclude the corresponding smaller polygons from the Euclidean allocation output, retaining the original larger polygons. This adjustment is based on the assumption that larger settlements have more comprehensive and digitized natural and man-made features, which facilitate their division into more manageable and distinct units.

## 2.4 Urban, Rural and Refugee Camp Classification

In constructing the sampling frame, it is crucial to distinguish between urban, rural, and refugee camp strata due to variations in population density and distribution. Additionally, the boundaries of refugee camps should be carefully respected, as they are specifically designated for refugee populations. The boundary of the refugee camp in Cameroon was initially provided by the UNHCR team. However, after overlaying the camp boundary onto high-resolution imagery and conducting a visual assessment, it was determined that the boundary should be adjusted to reflect recent changes and expansions within the camp. The modified camp boundary was shared with the UNHCR ground team in Cameroon, who reviewed and approved the proposed changes (figure 6).

In Cameroon, aside from grid-based data, there are no publicly available digitized administrative boundaries to differentiate between urban and rural strata. Consequently, the creation of urban and rural boundaries relied on internationally recognized standards, such as the Degrees of Urbanization framework (Dijkstra et al. 2021), along with additional techniques to account for recent urban expansion. To enhance the “Degree of Urbanization” method provided by EUROSTAT, the GHS Settlement Model grid (GHS-SMOD) classifies settlement typologies based on population size, population density, and built-up area density, categorizing areas into urban and rural classes (Florczyk et al. 2019).

The geographic locations of urban and rural strata in this study were delineated using the SMOD Level 1 dataset for 2025. This dataset provides three classes: two urban and one rural. The two urban classes were combined to define urban areas, while the remaining regions in Cameroon were categorized as rural. However, the SMOD urban classes often fall within the middle of densely populated areas, and their boundaries frequently intersect buildings and structures. Since the outlines of pre-enumeration areas and sample units

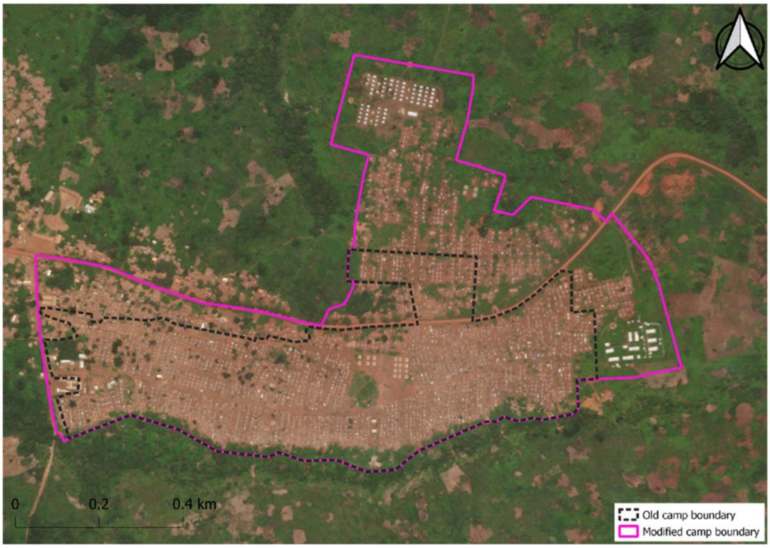


**Figure 6. Comparison of the original and modified refugee camp boundaries for Lolo.** The date of the original camp boundaries is unknown. The source of the satellite imagery is the ESRI basemap in QGIS.

should align with visible ground features or, at the very least, avoid cutting through structures, this classification proved inadequate for their construction. To address these issues, various geospatial techniques were employed:

(1) The settlement boundaries developed in Section 2.2.3 were used as the foundation for delineating urban and rural areas.
(2) Spatial aggregation techniques were employed to merge settlements within a 100-m radius, ensuring a more cohesive representation of urban and rural zones.
(3) Minimum bounding geometry approaches, specifically the generation of convex hulls around each settlement, were applied to refine the gridded outline of the polygonization process and ensure comprehensive coverage of the settled areas.
(4) Urban and rural strata were defined within the respective regions using the SMOD datasets.

These techniques reduced the risk of dividing contiguous communities or cutting through buildings in the urban and rural boundaries. Additionally, contiguous settlements were classified as either urban or rural, making the classification more applicable within the context of the sampling frame. In conjunction with the automated methods, manual adjustments were made where necessary. For example, if the classification indicated that more than 90 percent of a specific administrative boundary was urban, the entire boundary was classified as urban strata.

## 2.5 Implementing the preEA Tool

*2.5.1* Initial preEA production.

The output of the preEA tool was referred to as the pre-Enumeration Areas (preEAs) because it requires laboratory examination and adjustments before deployment in the field. The preEA tool is a Python-based QGIS plugin designed with a user-friendly interface, allowing individuals with limited GIS experience to utilize it effectively. Although still in development and not yet released, the tool's application and validation have been demonstrated across various contexts (Qader et al. 2021; 2022; 2023).

The tool utilizes gridded refugee population data alongside existing digitized boundaries (e.g. roads, railways, waterways) to divide the country into small, identifiable units (building blocks). These units are then aggregated according to user-defined criteria, such as population, geographic area, and other relevant constraints. The pre-EA boundaries were generated exclusively for arrondissements, administrative sub-division units with a minimum of 100 refugees, based on modelled ProGres data. Due to variations in refugee population density, spatial distribution, and stratum characteristics, different settings were applied to generate pre-EAs for each stratum.

Input datasets.

- Processed OpenStreetMap (OSM) line datasets (e.g. roads, railways, and waterways)
- Euclidean Allocation zones
- Uncrossable features
- Gridded refugee population data

Input parameters. The primary hard constraints were the estimated number of refugees, geographic area, and uncrossable borders. The maximum refugee population constraint was set at 100 for urban, rural, and camp settings. Additionally, the maximum geographic area for these settings was constrained to 2 $km^2$ for urban areas, 9 $km^2$ for rural areas, and 3 $km^2$ for camp settings.

*2.5.2 Post pre-EA modifications.*

When large geographic datasets were combined with multiple lines, borders, and constraints, polygons with undesirable or complex shapes may result. To address this, both automatic and manual adjustments were made to refine the pre-EA outcomes. Manual adjustments primarily focused on areas where digitized boundaries were limited, leading to the creation of pre-EAs with disproportionately large refugee populations. The pre-EA tool package also includes a tool that facilitates the automatic elimination of polygons that do not meet pre-EA criteria. Users can define the minimum refugee population size and geographic area for the pre-EAs that require merging. The tool automatically merges selected polygons with their neighbors, provided they do not cross the defined administrative boundaries and share the longest boundary.

Given the low population density and sparse distribution of refugees, only minimum geographic area constraints were applied in the tool. The following criteria were used for merging pre-EAs based on area:

- Urban: minimum area $\leq 6\ km^2$
- Rural: minimum area $\leq 10\ km^2$
- Refugee camp: minimum area $\leq 1\ km^2$

# 3. RESULT

## 3.1 Urban and Rural Classification

Based on the employed classification methods, 98 percent of Cameroon's land was designated as rural, while the combined total of urban areas and refugee camps accounted for less than 2 percent (figure 7). Of the refugees recorded in the UNHCR ProGres database, 46 percent resided in rural areas, with 31 percent in refugee camps and 23 percent in urban areas. The stratification approach used, which relies on the Global Human Settlement Layer (GHSL) dataset, applies a global definition to distinguish between urban and rural areas (Pesaresi et al. 2019). As a result, these classifications may not align with local definitions or alternative methods for achieving the same objective.

## 3.2 PreEA Boundary Outputs

The initial application of the pre-EA tool divided the refugee-populated area, encompassing urban, rural, and camp strata, into 157,318 building blocks (figure 8a). After the merging process was completed, the tool generated 25,582 pre-EAs across all strata (figure 8b). Following both automatic and manual adjustments, the pre-EA outputs were refined, reducing the total number of pre-EAs to 22,810 (Urban = 7,623; Rural = 14,114; Camps = 1,073).

In general, the pre-EA boundaries have met the criteria for defining enumeration areas. These criteria include adhering to significant uncrossable features, such as major roads and rivers, and ensuring that pre-EAs are nested within administrative boundaries (figure 9a and b). Additionally, the pre-EAs were aligned with visible features on the ground, particularly in urban areas, facilitating easier ground data collection and navigation (figure 9c). Moreover, the

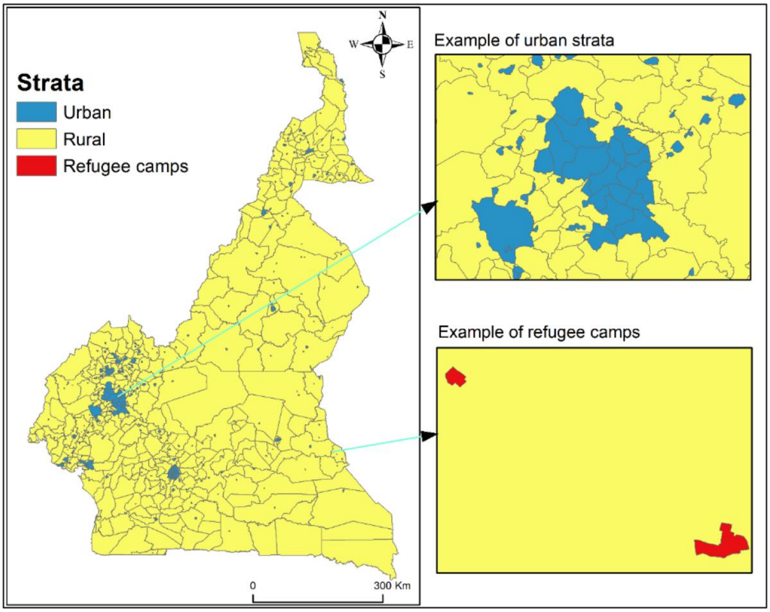


**Figure 7. Classification of urban, rural, and refugee camp strata in Cameroon for 2025.**

pre-EAs were sufficiently large to maintain the privacy of the households within them.

## 4. DISCUSSION

A key step in sampling refugee populations is creating a suitable sampling frame. This could be a refugee-only list or a general population list identifying displaced individuals, often sourced from government records or international organizations like UNHCR. However, such data pose challenges due to limited coverage, lack of detail, and absence of GPS coordinates. Traditional methods to build national sampling frames are also costly and time-consuming. These data gaps hinder effective planning and evaluation in low- and middle-income countries, highlighting the need for low-cost, innovative solutions. This study explores the preEA tool, which uses geospatial data and detailed refugee estimates to build customized sampling frames—particularly valuable in data-scarce, resource-limited settings.

This study faced several limitations, including the coarse spatial resolution of refugee datasets and incomplete or poor-quality digitized images of natural and man-made features in Cameroon. The accuracy of the gridded refugee population—though beyond this study's scope—directly affects the creation

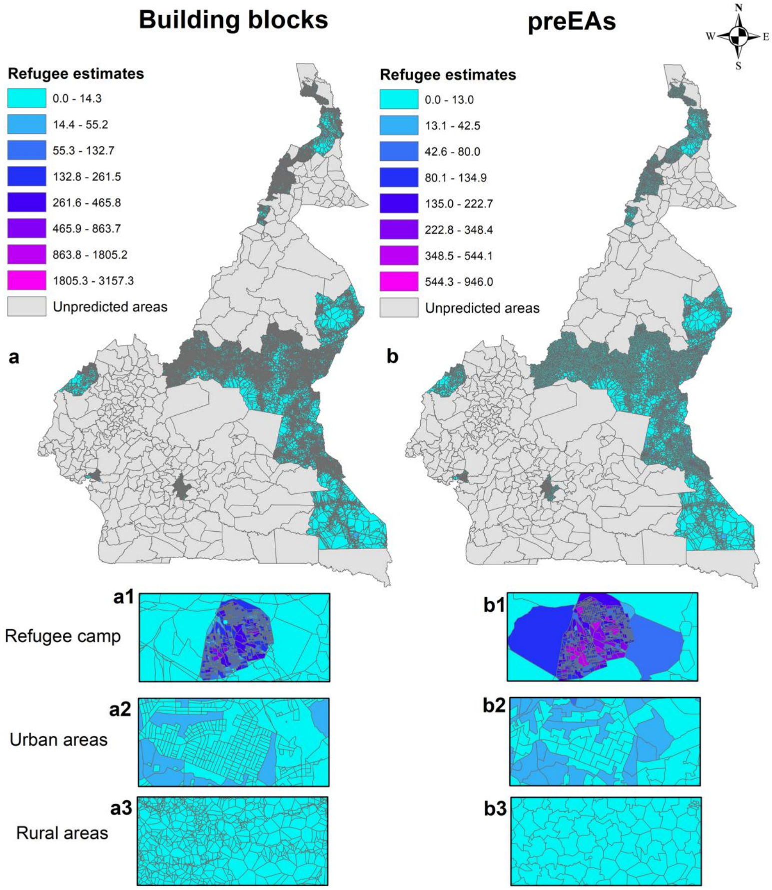


**Figure 8. pre-EA Tool outputs, including building blocks and pre-EAs across the predicted refugee areas in Cameroon.**

of sampling units, as it was a key input in defining their size. When interpreting the estimated refugee population within the national enumeration area-based sampling frame, limitations must be acknowledged. The UNHCR ProGres database, the primary data source, may not fully capture the refugee population in Cameroon. Outside camps, ProGres data was recorded at varying administrative levels, but digital boundaries do not exist beyond arrondissements (level 3). To reconcile ProGres data with georeferenced datasets, various spatial joining methods were applied. Additionally, multiple covariates were used to improve refugee population estimates, introducing uncertainties from model choice, data quality, and methodology. These uncertainties, in turn, affect the customized national sampling frame (see Darin et al. 2024, for details).

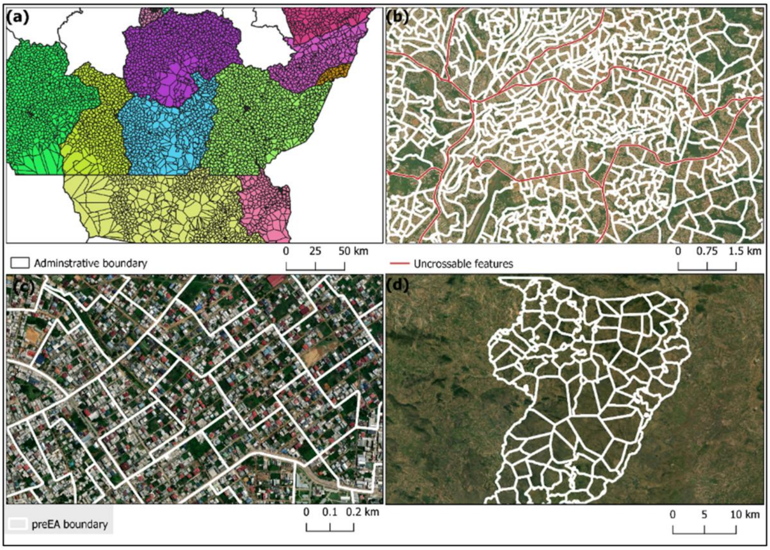


**Figure 9. Outline of the pre-EA boundaries overlaid on the ESRI high-resolution satellite imagery basemap.**

Regarding the limitation of digitized images of natural and man-made features, volunteer-digitized data like OSM often concentrates on urban areas, providing extensive coverage there, while rural and sparsely populated regions remain underrepresented or inaccurately mapped. Zielstra and Zipf (2010) found OSM data quality declined significantly from city centers to rural areas in Germany. Similarly, Borkowska and Pokonieczny (2022) noted the highest OSM data quality in Poland occurred in wealthier, urbanized areas. This urban bias is driven by shared public-private interests and the visibility of ground features, which are lacking in rural regions. In this study, digitized data in urban areas supported the creation of pre-enumeration areas (pre-EAs), with boundaries often matching visible features. In contrast, rural areas required different geospatial techniques to create smaller polygons, often with boundaries misaligned to ground features due to limited digitized data. The flexibility of the pre-EA tool was also restricted in some areas, where missing or poor-quality data led to irregularly shaped sampling units with high refugee populations, necessitating further manual adjustments.

The lack of complete attribute information and spatial coverage affected the automatic generation of the enumeration area-based sampling frame. Certain boundaries—such as major roads, rivers, and other obstacles—needed to be marked as uncrossable to aid surveyor navigation, based on available dataset attributes. However, missing or inaccurate data could misrepresent these features. Schott et al. (2024) found that while OSM attributes and geometry were generally accurate across 1,000 objects in Germany, land-use data was sometimes missing. Attribute quality also varies by location: only 9.35 percent of OSM roads in Baghdad had names (Gatea and Al-Bakri 2023), compared to 76 percent in England (Haklay 2010).

A major challenge in implementing geospatial solutions in low- and middle-income countries is the limited capacity of government sectors to adopt these methods due to skill gaps. Building this capacity through targeted training and support from international and local organizations is essential. Additionally, it is important to consider the sensitivity of the final results. Disclosure of the exact locations of vulnerable populations, such as refugees, could expose them to risks. As the statistical models used in this study primarily reveal the relative spatial distribution of refugees rather than exact population sizes, policies that negatively affect refugee populations could misuse such data. As outlined in Article 42 of the UNHCR Data Protection Policy (UNHCR, 2024b), it is advisable to implement institutional policies that protect refugees' rights by imposing access restrictions.

We recognize the difficulty in directly comparing costs and timelines between our approach and manual methods for delineating census EAs, as most countries do not disaggregate census expenditures to specifically identify the costs related to manual EA delineation. For instance, the Kenya National Bureau of Statistics (KNBS) received government and donor support to implement the 2019 Census, which had an estimated total cost of approximately Kshs. 18.5 billion (about USD 143 million) (Kenya National Bureau of Statistics (KNBS) 2019). Similarly, the Ghana Statistical Service (GSS), which is mandated to conduct national censuses, employed 75,000 Data Field Officers for their census operations (GhanaFact 2021). According to the finance minister, the exercise was projected to cost the government around 521 million cedis (approximately USD 36 million) (GhanaFact 2021). It is also worth noting that traditional survey processes can take two to three years to complete (Grosh and Glewwe 1995). If this work had been carried out manually, it would have required the recruitment of a team of cartographers, along with extensive training to enable them to delineate enumeration areas using high-resolution satellite imagery. This process would have involved the manual creation of thousands of enumeration boundaries, taking into account various constraints such as population density, geographic characteristics, and administrative requirements. Even after the initial training and production of the outputs, substantial effort would have been required to correct geometric inaccuracies inherent in manual digitization. These factors contribute to why such processes often span several years and demand significant financial resources. In contrast, the total cost for modeling the refugee population and developing the customized enumeration area-based sampling frame area was just USD 80,000. The entire methodological and data production effort was carried out by a two-member team. Another recent effective application of the pre-EA tool was demonstrated during the recent Household Budget Survey in Armenia, conducted by the World Bank (Ismailakhunova et al. 2025). In this case, the enumeration area-based sampling frame was successfully generated for Armenia in under three months using limited resources (Ismailakhunova et al. 2025).

With regard to an efficiency comparison between our approach and grid-based methods, we currently lack sufficient data and references to draw a definitive conclusion. However, a key distinction lies in the way each method delineates sampling units. Our approach generates boundaries that are aligned with visible ground features such as roads, infrastructure, and natural landmarks, which enhances the practicality and accuracy of field implementation. In contrast, grid-based methods often produce boundaries that do not correspond to physical features on the ground. This misalignment can lead to buildings being split across multiple grids, thereby introducing complications in weighting calculations and potential biases in population estimates.

Despite its limitations, this study presents a novel and cost-effective method for creating a digital enumeration area-based sampling frame tailored to specific subpopulations or the general population. As digitized boundaries improve and volunteered geographic data expands, this automated approach will face fewer constraints. For example, Iran's road network density increased notably from 2008 to 2016 (Minaei 2020), and OSM is estimated to be 83 percent complete globally, with full street networks in over 40 percent of countries (Barrington-Leigh and Millard-Ball 2017). Microsoft's Bing Maps has revealed 47.8 million km of roads, with only 1,165 km missing from OSM (Microsoft Road Detection 2024). Future research could also leverage more complete commercial road networks (Strano et al. 2017) and integrate additional data sources—such as call detail records, satellite imagery, and smartphone traces—to better capture refugee dynamics. In light of global funding constraints, NSOs and international organizations increasingly seek cost-effective tools like the pre-EA Tool to reduce costs in census and survey planning.

## 5. CONCLUSION

The national sampling frame is a list of PSUs, such as enumeration areas from a recent census, that are often used for household surveys. Inaccurate and outdated sample frames can lead to biased survey results due to coverage errors. This may be more evident when surveying populations that are hard-to-reach or forcibly displaced since they are not taken into consideration when developing a common census enumeration area-based sampling frame. Cameroon's last census was carried out in 2005, and no data are readily available to generate enumeration areas for a representative refugee survey sampling frame. As a pilot project, we partnered with UNHCR to investigate the potential of a recently developed pre-EA tool and make use of available geospatial datasets to create an enumeration area-based sampling frame that is specifically tailored for the refugee population in Cameroon. For the first time, the work produced publicly accessible digitized urban and rural boundaries in Cameroon that are in line with administrative borders and do not follow a grid system. According to our stratification, out of all the refugees listed in the UNHCR ProGres database, 46 percent live in rural areas, while 31 percent and 23 percent are situated in camps and urban areas, respectively. The outcome showed that the strategy was successful in creating the customized national sample frame in a timely, cost-effective, and limited-resource-wise manner. In addition, the generated sampling frame satisfies surveyor expectations as well as international standards. The method can be adopted to create a customized, enumeration area-based sampling frame in other countries due to its effectiveness, reliability, and flexibility.

## DATA AVAILABILITY

The generated customized enumeration area-based sampling frame can be publicly available, but restrictions apply to the availability of refugee population estimates within the sampling units, as these data are not publicly accessible. However, upon reasonable request and with approval from the UN High Commission for Refugees, the enumeration area-based sampling frame with the estimated number of refugees can be available from the authors.

## REFERENCES

Barrington-Leigh, C., and Millard-Ball, A. (2017), “The World’s User-Generated Road Map is More than 80% Complete,” *PLoS ONE*, 12(8), e0180698. https://doi.org/10.1371/journal.pone.0180698

Blankespoor, B., Kilic, T., Murray, S., Prydz E. B., and Wild M. (2018), Can remote sensing data complement or even replace the current sampling frames in household surveys in developing countries? https://ui.adsabs.harvard.edu/abs/2018AGUFMIN33B0844B

Bondarenko, M., Kerr, D., Sorichetta A., and Tatem A. J., (2020), *Census/Projection-Disaggregated Gridded Population Datasets for 51 Countries across Sub-Saharan Africa in 2020 Using Building Footprints*. WorldPop, University of Southampton, UK. https://doi.org/10.5258/SOTON/WP00682

Boo, G., Darin, E., Thomson, D. R., and Tatem, A. J. (2020), “A Grid-Based Sample Design Framework for Household Surveys [Version 1; Peer Review: 2 Approved, 1 Approved with Reservations],” *Gates Open Res*, 4, 13. https://doi.org/10.12688/gatesopenres.13107.1

Borkowska, S., and Pokonieczny, K. (2022), “Analysis of OpenStreetMap Data Quality for Selected Counties in Poland in Terms of Sustainable Development,” *Sustainability*, 14, 3728. https://doi.org/10.3390/su14073728

Cajka, J., Amer, S., Ridenhour, J., and Allpress, J. (2018), “Geo-Sampling in Developing Nations,” *International Journal of Social Research Methodology*, 21, 729–746. https://doi.org/10.1080/13645579.2018.1484989

Darin, E., Dicko, A. H., Galal, H., Jimenez, R. M., Park, H., Tatem, A. J., and Qader, S. (2024), “Mapping Refugee Populations at High Resolution by Unlocking Humanitarian Administrative Data,” *Journal of International Humanitarian Action*, 9, 14. https://doi.org/10.1186/s41018-024-00157-6

Department of Economic and Social Affairs Statistics Division (UNSD) (2017), *Principles and Recommendations for Population and Housing Censuses*, New York. https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Principles_and_Recommendations/Population-and-Housing-Censuses/Series_M67rev3-E.pdf

Dijkstra, L., Florczyk, A. J., Freire, S., Kemper, T., Melchiorri, M., Pesaresi, M., and Schiavina, M. (2021), "Applying the Degree of Urbanisation to the Globe: A New Harmonised Definition Reveals a Different Picture of Global Urbanisation," *Journal of Urban Economics*, 125, 103312. https://doi.org/10.1016/j.jue.2020.103312

Dooley, C. A., Boo, G., Leasure D. R., and Tatem A. J., (2020), *Gridded Maps of Building Patterns throughout Sub-Saharan Africa, Version 1.1*, University of Southampton: Southampton, UK. Source of building footprints "Ecopia Vector Maps Powered by Maxar Satellite Imagery"© 2020. https://doi.org/10.5258/SOTON/WP00677

Eckman, S., and Himelein, K. (2020), Methods of Geo-Spatial Sampling. In *Data Collection in Fragile States: Innovations from Africa and Beyond*, J. Hoogeveen & U. Pape eds., Washington, DC: Springer International Publishing. pp. 103–128. https://doi.org/10.1007/978-3-030-25120-8_7

——— (2022), Innovative Sample Designs for Studies of Refugees and Internally Displaced Persons. In *Migration Research in a Digitised World: Using Innovative Technology to Tackle Methodological Challenges*, S. Pötzschke & S. Rinken eds., Springer International Publishing. pp. 15–34. https://doi.org/10.1007/978-3-031-01319-5_2

Ecopia.AI and Maxar Technologies (2020), Digitise Africa data.

ESRI 2025 (2023), Euclidean Allocation (Spatial Analyst). Software: ArcGIS Pro. 3.1.3; https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/euclidean-allocation.htm.

Florczyk, A. J., Corbane, C., Ehrlich, D., Freire, S., Kemper, T., Maffenini, L., Melchiorri, M., Pesaresi, M., Politis, P., Schiavina, M., Sabo, F., and Zanchetta, L. (2019), *GHSL Data Package 2019*, Luxembourg: Publications Office of the European Union. https://doi.org/10.2760/0726

Gallego, F. J., Kerdiles, H., and Leo, O. (2015), Using GIS and Remote Sensing to Build Master Sampling Frames for Agricultural Statistics. EUR 27597 EN. Publications Office of the European Union. https://doi.org/10.2788/438056

Galway, L. P., Bell, N., Sae, A. S., Hagopian, A., Burnham, G., Flaxman, A., Weiss, W. M., Rajaratnam, J., and Takaro, T. K. (2012), "A Two-Stage Cluster Sampling Method Using Gridded Population Data, a GIS, and Google EarthTM Imagery in a Population-Based Mortality Survey in Iraq," *International Journal of Health Geographics*, 11, 12. https://doi.org/10.1186/1476-072X-11-12

Gatea, K. Z., and Al-Bakri, M. (2023), "Measuring the Attribute Accuracy and Completeness for the OpenStreetMap Roads Networks for Two Regions in Iraq," *Journal of Engineering*, 29, 156–168. https://doi.org/10.31026/j.eng.2023.05.12

GhanaFact (2021), FACTSHEET: What you need to know about the 2021 Population and Housing Census. https://ghanafact.com/2021/06/factsheet-what-you-need-to-know-about-the-2021-population-and-housing-census-phc/#:~:text=With%20the%20mandate%20to%20conduct, the%20government%20521%20million%20cedis

Grosh, M. E., and Glewwe, P. (1995), "A Guide to Living Standards Measurement Study Surveys and their Data Sets." LSMS Working Paper No 120. https://documents1.worldbank.org/curated/en/270551468764720584/pdf/multi-page.pdf.

Haklay, M. (2010), "How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets," *Environment and Planning B: Planning and Design*, 37, 682–703. https://doi.org/10.1068/b35097

Himelein, K., Eckman, S., and Murray, S. (2014), "Sampling Nomads: A New Technique for Remote, Hard-to-Reach, and Mobile Populations," *Journal of Official Statistics*, 30, 191–213. https://doi.org/10.2478/jos-2014-0013

International Peace Institute (IPI) (2018), Reaching Internally Displaced Persons to Achieve the 2030 Agenda. https://www.ipinst.org/wp-content/uploads/2018/11/1811_IDPs-and-SDGs.pdf

Ismailakhunova, S., Purevjav, A.-O., Byambasuren, T., and Qader, S. (2025), Exploring Geospatial-Based Approaches to Develop a Pre-Census National Sampling Frame in Armenia.

Policy Research Working Paper; 11039. © World Bank. http://hdl.handle.net/10986/42698 License: CC BY 3.0 IGO

Kenya National Bureau of Statistics (KNBS) (2019), 2019 census backgrounder. https://www.knbs.or.ke/2019-census-backgrounder/#:~:text=THE%20CENSUS%20PROCESS,18.5%20billion

Martin-Shields, C. P., Camacho, S., Taborda, R., and Ruhe, C. (2019), "Digitalisation in the lives of urban migrants: Evidence from Bogota," Discussion Paper, No. 12/2019, ISBN 978-3-96021-103-7, Bonn, Germany: Deutsches Institut für Entwicklungspolitik (DIE). https://doi.org/10.23661/dp12.2019

Mather, M., and Cortes, R. (2016), Demographic Data: Censuses, Registers, Surveys. In The Blackwell Encyclopedia of Sociology, G. Ritzer, ed. https://doi.org/10.1002/9781405165518.wbeosd018.pub2 (Accessed 12/03/2024)

Microsoft Road Detection (2024), Mined Roads: https://gee-community-catalog.org/projects/msroads/

Miller, A. C., Rohloff, P., Blake, A., Dhaenens, E., Shaw, L., Tuiz, E., Grandesso, F., Mendoza Montano, C., and Thomson, D. R. (2020), "Feasibility of Satellite Image and GIS Sampling for Population Representative Surveys: A Case Study from Rural Guatemala," *International Journal of Health Geographics*, 19, 56. https://doi.org/10.1186/s12942-020-00250-0

Minaei, M. (2020), "Evolution, Density and Completeness of OpenStreetMap Road Networks in Developing Countries: The Case of Iran," *Applied Geography*, 119, 102246. https://doi.org/10.1016/j.apgeog.2020.102246

Office for National Statistics (ONS), released 2 November (2022), ONS website, methodology, Quality and methodology information (QMI) for Census 2021. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/methodologies/qualityandmethodologyinformationqmiforcensus2021

Office for the Coordination of Humanitarian Affairs (OCHA) (2023), Cameroon situation report. https://reports.unocha.org/en/country/cameroon/

Pape, U. J., and Wollburg, P. R. (2019), *Estimation of Poverty in Somalia Using Innovative Methodologies*, Policy Research working paper; no. WPS 8735. Washington, D.C.: World Bank Group.

Parsaeian, M., Mahdavi, M., Saadati, M., Mehdipour, P., Sheidaei, A., Khatibzadeh, S., Farzadfar, F., and Shahraz, S. (2021), "Introducing an Efficient Sampling Method for National Surveys with Limited Sample Sizes: Application to a National Study to Determine Quality and Cost of Healthcare," *BMC Public Health*, 21, 1414. https://doi.org/10.1186/s12889-021-11441-0

Pesaresi, M., Florczyk, A., Schiavina, M., Melchiorri, M., and Maffenini, L. (2019), GHS settlement grid, updated and refined REGIO model 2014 in application to GHS-BUILT R2018A and GHS-POP R2019A, multitemporal (1975-1990-2000-2015), R2019A. European Commission, Joint Research Centre (JRC) https://doi.org/10.2905/42E8BE89-54FF-464E-BE7B-BF9E64DA5218PID: http://data.europa.eu/89h/42e8be89-54ff-464e-be7b-bf9e64da5218

Qader, S., Lefebvre, V., Tatem, A., Pape, U., Himelein, K., Ninneman, A., Bengtsson, L., and Bird, T. (2021), "Semi-Automatic Mapping of Pre-Census Enumeration Areas and Population Sampling Frames," *Humanities and Social Sciences Communications*, 8, 3. https://doi.org/10.1057/s41599-020-00670-0

Qader, S. H., Lefebvre, V., Tatem, A. J., Pape, U., Jochem, W., Himelein, K., Ninneman, A., Wolburg, P., Nunez-Chaim, G., Bengtsson, L., and Bird, T. (2020), "Using Gridded Population and Quadtree Sampling Units to Support Survey Sample Design in Low-Income Settings," *International Journal of Health Geographics*, 19, 10. https://doi.org/10.1186/s12942-020-00205-5

Qader, S. H., Harfoot, A., Kuepie, M., Darin, E., Juran, S., Lazar, A., and Tatem, A. (2022), *"National Automatic preEnumeration Areas (preEAs) in Burkina Faso (2019), Version 1.0."*, WorldPop, University of Southampton. https://data.worldpop.org/repo/wopr/BFA/preEAs/v1.0/

Qader, S. H., Chamberlain, H., Kuepie, M., Hunt, F. K., Lazar, A., and Tatem, A. (2023), *"Field Testing of Pre-Enumeration Areas Created Using Semi-Automated Delineation Approach, Democratic Republic of Congo."*, WorldPop, University of Southampton. https://dx.doi.org/10.5258/SOTON/WP00759.

Reichel, D., and Morales, L. (2017), "Surveying Immigrants without Sampling Frames—Evaluating the Success of Alternative Field Methods," *CMS*, 5, 1. https://doi.org/10.1186/s40878-016-0044-9

Runfola, D., Anderson, A., Baier, H., Crittenden, M., Dowker, E., Fuhrig, S., Goodman, S., Grimsley, G., Layko, R., Melville, G., Mulder, M., Oberman, R., Panganiban, J., Peck, A., Seitz, L., Shea, S., Slevin, H., Youngerman, R., and Hobbs, L. (2020), "geoBoundaries: A Global Database of Political Administrative Boundaries," *PLoS ONE*, 15, e0231866. https://doi.org/10.1371/journal.pone.0231866.

Schott, M., Zell, A., Lautenbach, S., Sumbul, G., Schultz, M., Zipf, A., & Demir, B. (2024). Analyzing and Improving the Quality and Fitness for Purpose of OpenStreetMap as Labels in Remote Sensing Applications. In D. Burghardt, E. Demidova, & D. A. Keim (Eds.), *Volunteered Geographic Information: Interpretation, Visualization and Social Context* (pp. 21-42). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-35374-1_2

Strano, E., Giometto, A., Shai, S., Bertuzzo, E., Mucha, P. J., and Rinaldo, A. (2017), "The Scaling Structure of the Global Road Network," *Royal Society Open Science*, 4, 170590. https://doi.org/10.1098/rsos.170590

Thomson, D. R., Rhoda, D. A., Tatem, A. J., and Castro, M. C. (2020), "Gridded Population Survey Sampling: A Systematic Scoping Review of the Field and Strategic Research Agenda," *International Journal of Health Geographics*, 19, 34. https://doi.org/10.1186/s12942-020-00230-4

Thomson, D. R., Stevens, F. R., Ruktanonchai, N. W., Tatem, A. J., and Castro, M. C. (2017), "GridSample: An R Package to Generate Household Survey Primary Sampling Units (PSUs) from Gridded Population Data," *International Journal of Health Geographics*, 16, 25. https://doi.org/10.1186/s12942-017-0098-4

United Nations High Commissioner for Refugees (UNHCR) (2023), Global Trends 2022. https://www.unhcr.org/global-trends-report-2022

——— (2024a), Forced Displacement Survey (FDS https://www.unhcr.org/what-we-do/reports-and-publications/data-and-statistics/forced-displacement-survey

——— (2024b), General Policy on Personal Data Protection and Privacy. https://www.refworld.org/docid/63d3bdf94.html

Valliant, R., Dever, J., and Kreuter, F. (2018), Practical Tools for Designing and Weighting Survey Samples. https://doi.org/10.1007/978-1-4614-6449-5

Wagenaar, B. H., Augusto, O., Ásbjörnsdóttir, K., Akullian, A., Manaca, N., Chale, F., Muanido, A., Covele, A., Michel, C., Gimbel, S., Radford, T., Girardot, B., Sherr, K., Manuel, J. L., Hicks, L., Mahumane, A., Pfeiffer, J., Gloyd, S., . . . Cuembelo, F, with input from the, I. S. T (2018), "Developing a Representative Community Health Survey Sampling Frame Using Open-Source Remote Satellite Imagery in Mozambique," *International Journal of Health Geographics*, 17, 37. https://doi.org/10.1186/s12942-018-0158-4

Willis, G. B., Smith, T. W., Shariff-Marco, S., and English, N. (2014), "Overview of the Special Issue on Surveying the Hard-to-Reach," *Journal of Official Statistics*, 30, 171–176. https://doi.org/10.2478/jos-2014-0011

Zielstra, D., and Zipf, A. (2010), "Quantitative Studies on the Data Quality of OpenStreetMap in Germany," in *Proceedings of GIScience*, Vol. 2010, no. 3.


Journal of Survey Statistics and Methodology, 2026, 00, 1–25
https://doi.org/10.1093/jssam/smaf027
Article