Photo Credit: UN Women/Srijana Bhatta

TOWARDS CARING CITIES: A GEOSPATIAL ANALYSIS IN VIET NAM

Methodological Note

DATA-POP ALLIANCE

THÀNH PHỐ ĐÀ NẴNG

UN WOMEN

FOR ALL WOMEN AND GIRLS

TABLE OF CONTENTS

I. Introduction

This methodological note outlines the technical approach used for the Geospatial Analysis of Care Systems in Da Nang, Viet Nam, conducted by UN Women in collaboration with the Da Nang Women's Union and Data-Pop Alliance. The analysis is intended to generate actionable evidence to direct investments towards transforming care systems at the local level by identifying underserved areas and populations with unmet care needs. Building on the methodology proposed by UNDP[1], the study advances innovative methods for mapping care systems by incorporating contextual factors, such as flooding, cyclone exposure and land-use patterns, that directly influence both care demand and service accessibility.

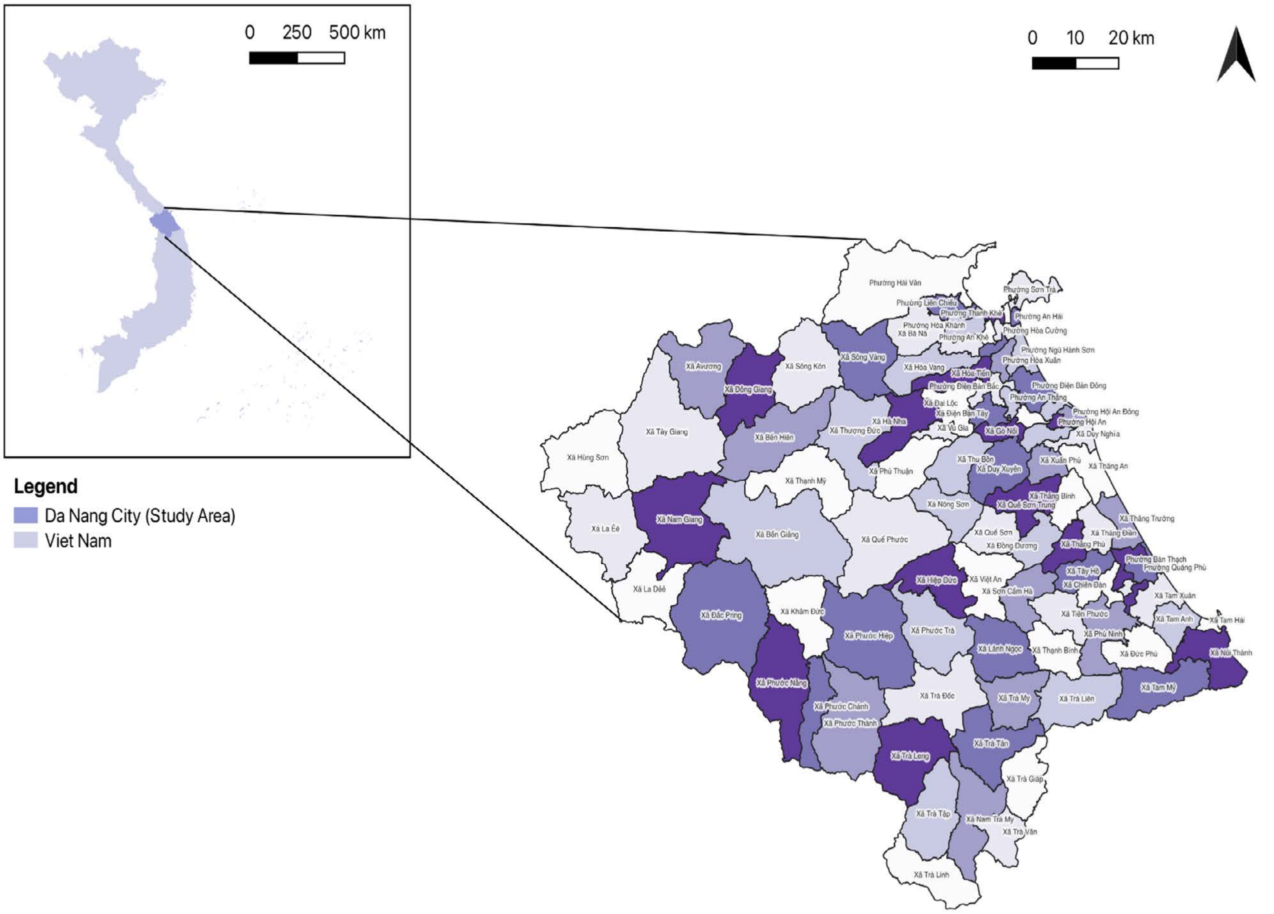
In Da Nang, communes represent the smallest administrative units responsible for localized governance, making them an appropriate scale for identifying spatial disparities in care provision. The analysis therefore examines both care demand and supply at the commune level, using an accessibility-based approach to detect potential discrepancies between where care-dependent populations live and where services are located. The demand-side modelling integrates demographic and socioeconomic characteristics, including age and sex distribution, alongside contextual variables such as income levels, land-use categories and climate-related risks. This multidimensional approach demonstrates Geographic Information Systems (GIS) added value in revealing inequities that remain hidden in traditional datasets, providing a more holistic understanding of Viet Nam's care ecosystem and informing more precise, needs-based interventions.

All geospatial datasets were processed using the EPSG:32649 coordinate reference system (WGS 84 / UTM Zone 49N), with metres as the unit of measurement. Administrative boundary shapefiles for the city and communes served as the basis for all spatial analyses.

The sections that follow describe the step-by-step procedures used to map care demand, integrate contextual factors and assess care services' spatial distribution. A dedicated section outlines administrative boundary data processing, and the methodological note concludes with a discussion of limitations and mitigation strategies.

1 UNDP Policy Note. 14. 2022. Mapping Care: Innovative tools for georeferencing care supply and demand in Latin America and the Caribbean.

Map 1: Study Area



Source: Authors' elaboration using Da Nang City Administrative Boundaries (2025)

II. Identifying Main Analytical Areas: Mixed-Methods Clustering

The identification of distinct geographic areas for the analysis is carried out through ADM3-level (commune) clustering using a mixed-methods approach. This methodology combines rigorous quantitative statistical techniques with contextual qualitative validation and rule-based adjustments to ensure that the resulting clusters accurately reflect the study area's socioeconomic, demographic and geographical dynamics. By integrating these complementary methods, the approach captures both statistically significant patterns and locally relevant nuances that may not be fully visible through quantitative analysis alone. The clustering process consists of three sequential components: quantitative cluster creation derived from standardized indicators; qualitative validation, during which local contextual knowledge and spatial interpretation are applied; and mixed-methods final definition, where quantitative outputs and qualitative insights are synthesized to generate robust, policy-relevant geographic clusters.

2.1. QUANTITATIVE CLUSTERING

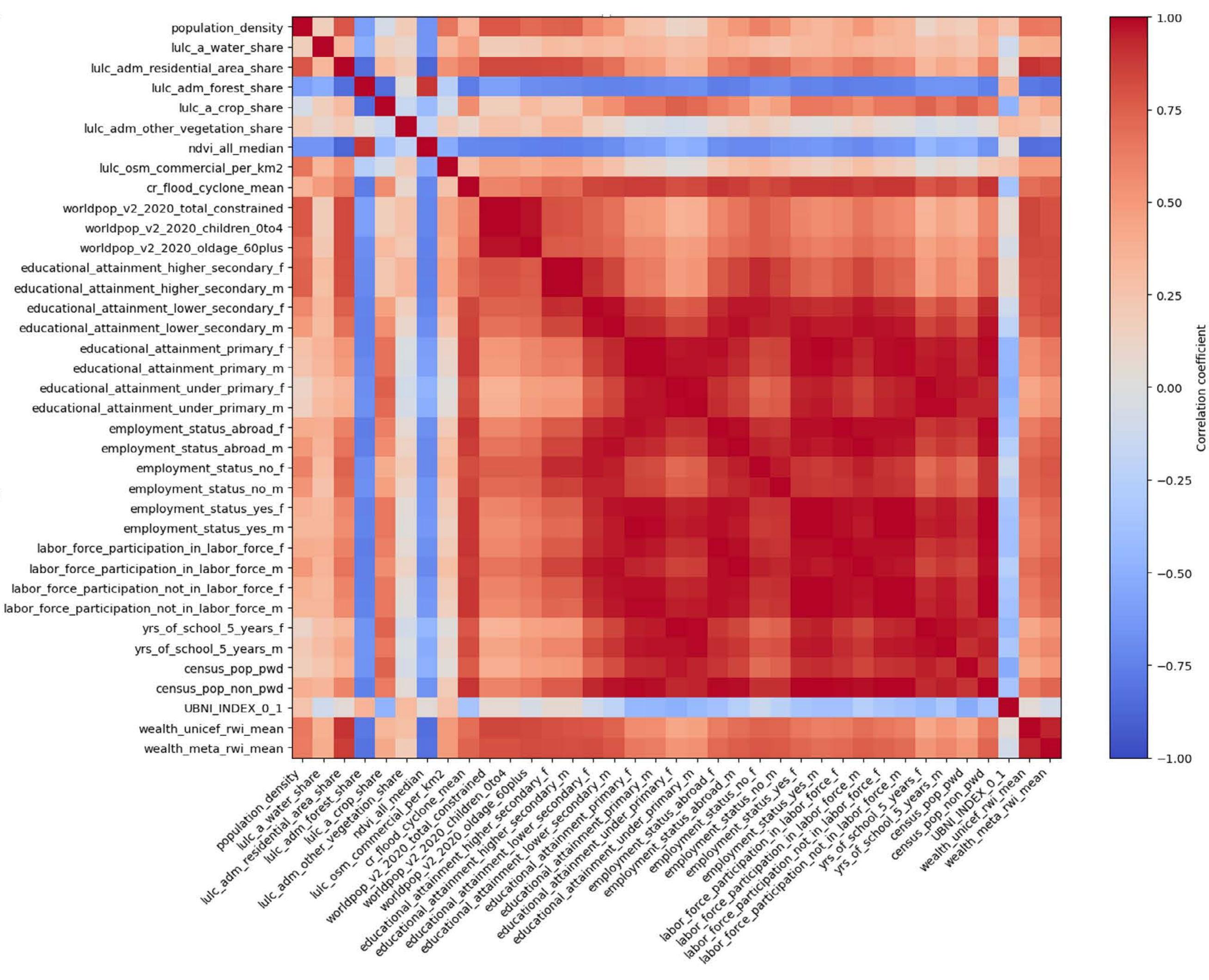
The clustering methodology's quantitative component applies hierarchical clustering at the ADM3 level. This approach constructs a hierarchical structure among administrative units by calculating distances based on a wide range of sociodemographic, economic and geographic characteristics. To ensure clusters reflect territorial realities, geographic distances between ADM3 centroids are incorporated into the distance matrix, reinforcing spatial coherence and avoiding disjointed or impractical geographic groupings.

Quantitative clustering aims to group administrative areas according to factors that are critical for nuanced spatial planning and policy analysis. The methodology seeks to achieve geographic contiguity that facilitates effective local administration, bring together areas with similar sociodemographic profiles and cluster communes sharing land-use characteristics relevant to urban development patterns. The approach also incorporates geographic features, particularly climate-related risks, and reflects socioeconomic integration by accounting for spatial variations in wealth and employment dynamics.

The analysis considers a comprehensive indicator set across several domains (Figure 1, Input Correlation Heatmap). Three main categories of quantitative inputs inform clustering: one for quantitative clustering and two for qualitative checks. The first category comprises baseline indicators, including administrative land-use classifications, commercial establishment distribution from OpenStreetMap (OSM), the Normalized Difference Vegetation Index, combined flood and cyclone risk levels, population density and WorldPop estimates of total, child and elderly populations. The second category integrates socioeconomic indicators, notably the

Meta Relative Wealth Index and UNICEF's Relative Wealth Index, alongside the Unsatisfied Basic Needs Index calculated from the 2019 census. The third category incorporates census-based variables estimated at the ADM3 level using pre-merger ADM2 census data, including educational attainment by sex, employment status and labour force participation by sex, years of schooling and population counts for persons with disabilities and non-disabled individuals. Together, these indicators provide a robust, multidimensional foundation for quantitatively identifying meaningful spatial clusters that reflect Da Nang's demographic, socioeconomic and environmental characteristics.

Figure 1: Input Correlation Heatmap



Source: Authors' elaboration using Da Nang City Administrative Boundaries (2025), WorldPop (2024), Administrative Land Use, IPUMS (2019), National Center for Hydrometeorological Forecasting (2023), Humanitarian OpenStreetMap (2025), Meta Data for Good (2021) and ESA Copernicus (2024)

2. 2. QUALITATIVE CLUSTERING

Following initial quantitative cluster schematic generation, typically produced by testing three-, four, five- or six-cluster configurations using different input-variable combinations, results undergo a qualitative verification stage. This validation step ensures that statistically derived groupings meaningfully correspond to contextual realities and insights documented in qualitative literature on Da Nang and Quang Nam's socioeconomic and geographic zones. Quantitative models alone cannot fully account for place-specific dynamics such as localized development patterns, administrative considerations, historical settlement structures or emerging climate vulnerabilities, making qualitative review indispensable.

During manual verification, cluster boundaries are examined in relation to local knowledge, expert input and spatial interpretation. While the statistical models' underlying structure is generally retained, adjustments are made where necessary to correct misalignments, reconcile boundary inconsistencies and incorporate contextual understanding not captured through numerical indicators alone. This iterative refinement ensures final clusters constitute coherent demographic and geographic units, strengthening relevance for care-needs analysis, vulnerability identification and spatially informed policy design.

2.3. MIXED-METHODS FINAL CLUSTER DEFINITION AND RATIONALE

The clustering structure's final definition applies a mixed-methods rationale to delineate four distinct clusters balancing statistical coherence with contextual relevance. This synthesis enables resulting geographic units to reflect demographic, socioeconomic and environmental characteristics most significant for interpreting care needs and spatial inequalities. Cluster assignments are grounded in clear quantitative thresholds, primarily related to population characteristics and land-use composition, while also incorporating auxiliary indices for wealth, climate risk and administrative land-use patterns to refine and validate categorization. The four primary clusters emerging from this process are the Urban Core, Coastal Urban Fringe, Rural Production Zone and Forest and Conservation Zone.

The **Urban Core** is defined principally by high population density, with administrative units classified as urban when they meet or exceed a threshold of 1,500 persons per square kilometre. This threshold aligns with established international guidelines, including Eurostat's criteria (2021) for determining urbanization degree using gridded population data.[2] To reinforce classification robustness, supporting datasets such as the Meta Relative Wealth Index and Administrative Land Use maps ensure designated urban areas also exhibit wealth, infrastructure and land-use characteristics typically associated with dense metropolitan environments.

The **Coastal Urban Fringe** represents the transitional zone between the highly concentrated Urban Core and more spatially extensive rural areas. This cluster is characterized by a lower population density threshold of 500 persons per square kilometre, capturing settlements maintaining partial urban characteristics while remaining influenced by surrounding peri-urban and coastal dynamics. Category refinement draws on multiple geospatial

2 Vietnam's Decree 42/2009/ND-CP defines urban centres using several criteria, including population size, population density in the inner area (generally from 4,000 to 10,000 persons per square kilometre depending on the grade of the urban centre), the proportion of non-agricultural labour, and the presence of required infrastructure and urban functions. These criteria are intended for classifying entire cities and towns rather than determining the urban or rural status of smaller administrative units such as communes. Because the available data do not include the necessary labour force or infrastructure variables, and because applying city-level density thresholds to commune-level units would not align with the purpose of the decree, this classification was not used in the analysis.

layers, including the Normalized Difference Vegetation Index, commercial and service facility distribution and flood and cyclone risk level classification. These factors help identify areas where climate hazard exposure intersects with coastal development and service accessibility patterns.

The broader rural territory is divided into two clusters, namely the **Rural Production Zone and Forest and Conservation Zone**, based on agricultural land share. A 10 per cent agricultural land threshold differentiates areas with high agricultural activity from those dominated by forest cover, conservation land or non-agricultural natural landscapes. This distinction reflects functional land-use differences with implications for livelihoods, mobility, service access and care needs. Classification is further refined using industrial and mining maps, the Meta Relative Wealth Index, Administrative Land Use data and the census-derived Unsatisfied Basic Needs Index. These supporting datasets ensure both rural clusters accurately capture socioeconomic disparities and the spatial interplay between productive land uses and environmental conservation zones.

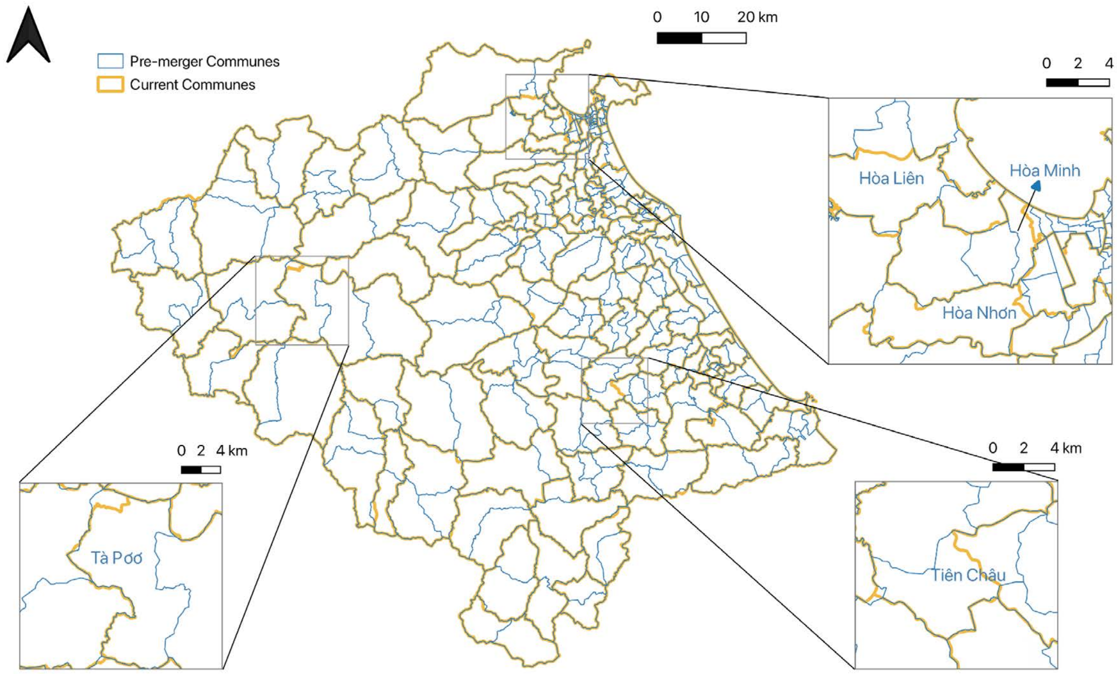
Through this integrated approach, the final clustering framework provides a nuanced, policy-relevant spatial typology that supports deeper analysis of care demand, vulnerability patterns and context-specific intervention design.

2.4. IDENTIFIED DISTORTED AREAS AND IMPLICATIONS

Da Nang City and the adjacent Quang Nam Province have undergone recent administrative boundary adjustments, particularly involving the reorganization and merger of commune-level administrative units (ADM3). These changes reflect ongoing government efforts to streamline governance structures; however, they introduce important methodological implications for geospatial analysis, especially when linking pre-merger demographic, socioeconomic and contextual indicators to newly defined administrative units. Because many datasets, including census information, climate risk layers and administrative records, were produced under earlier boundary configurations, careful spatial harmonization is required to avoid misalignment and misinterpretation.

Using the official Administrative Boundary Dataset (2025) as the reference geometry, together with historical boundary data compiled from UN OCHA (2025), a polygon-based comparison was conducted to quantify spatial changes. New commune boundaries (Map 2, in orange) were overlaid with their corresponding pre-merger counterparts (in blue) to visualize shifts in shape and extent. Pre-merger communes were then grouped to align with updated administrative units based on the criterion of maximum spatial overlap, ensuring that historical indicators could be aggregated consistently into the new configuration.

Map 2: Commune Boundary Comparison



Sources: Authors' elaboration using UN OCHA (2025) and Da Nang City Administrative Boundaries (2025)

This comparison revealed that several merged communes exhibit irregular or "distorted" geometries after alignment. These distortions typically occur in transition zones where multiple small polygons have been consolidated into a larger unit with non-contiguous or elongated shapes. Such artefacts are not merely geometric curiosities; they have direct methodological implications. Distorted areas are more prone to data allocation inaccuracies, particularly when redistributing pre-merger indicators such as population counts, land-use proportions or risk classifications. They may also influence distance-based measures, centroid calculations and spatial clustering outputs, underscoring the need for careful interpretation of results and transparent documentation of boundary harmonization procedures.

To assess the implications of administrative boundary changes for the geospatial analysis, the areas affected by ADM3 mergers were classified into three groups according to the spatial extent of the boundary modification and the demographic significance of the affected territory. This typology provides a clearer understanding of how boundary adjustments may influence indicator allocation, facility counts and subsequent analytical outputs. Table 1 presents more details on the classification criteria and affected communes.

1. **Group 1 – Minor to Moderate Impact:** This group includes areas where boundary shifts occurred primarily in moderately populated rural zones. Although population densities change only slightly, aggregated indicators such as total population counts are moderately affected due to population reassignment from pre-merger to post-merger units. These areas require careful realignment but do not substantially alter spatial patterns at the city or cluster level.

2. **Group 2 – Moderate Impact:** This group comprises densely populated or urbanized areas where pre-merger polygons were relatively small. Boundary adjustments in these locations exert a pronounced influence on aggregated indicators and significantly

affect location-based facility counts, including health and care centres. Even minor geometric shifts in dense urban zones can alter service-availability and accessibility spatial representation, making accurate boundary harmonization particularly critical.

3. **Group 3 – Minor Impact:** This group contains polygons that are predominantly forested or sparsely populated. Changes in these areas have negligible effects on density-based indicators and minimal influence on count-based measures. Although these units must still be aligned for consistency, their contribution to overall data distortion is limited.

Through this grouping approach, the analysis systematically identifies where administrative mergers may introduce greater uncertainty, enhancing transparency in indicator-result interpretation and spatial comparisons.

Table 1: Identified Distorted Areas and Implications

Previous Affected Commune	Current Affected Commune	Details of Administrative Reorganization	Area Characteristics	Impact Severity	Indicator Effect
Tiên Châu	Xã Sơn Cẩm Hà, Xã Tiên Phước	**Tiên Châu** is absorbed partially, with the larger section falling within **Xã Sơn Cẩm Hà**, and the remaining portion integrated into **Xã Tiên Phước**.	Contains areas of dense population (WorldPop Global 2 2020 visualization shown).	**Minor to Moderate**	**Minor effect on rate-based indicators** (e.g., population density) but a **moderate effect on aggregated indicators** (e.g., total population).
Hòa Liên, Hòa Minh, Hòa Nhơn	Phường Hải Vân, Phường Liên Chiểu, Phường Hòa Khánh, Phường Thanh Khê, Phường An Khê, Xã Bà Nà	This coastal/urban area saw extensive reconfiguration. The northern area of **Hòa Liên** was merged into **Phường Hải Vân**, and the rest into **Phường Liên Chiểu**. An eastern strip of Hòa Minh was combined with **Phường Thanh Khê**, while the major remainder stayed in **Phường Hòa Khánh**. Similarly, an eastern strip of **Hòa Nhơn** was added to **Phường An Khê**, with the rest placed in **Xã Bà Nà**.	Highly populated urban area, where boundary shifts involved small polygons.	**Moderate**	Significant shifts in **aggregated indicators** (e.g., total population) and specifically in the **counts of care facilities.**
Tà Pơơ	Xã Nam Giang, Xã Bến Giàng	The northwest corner of **Tà Pơơ** was cut to form **Xã Nam Giang**, while the remaining larger portion became part of **Xã Bến Giàng**.	Predominantly forested area with low population density.	**Minor**	Minimal overall impact on indicators.

To evaluate administrative mergers' influence on spatially derived indicators, population-based metrics were recalculated using the WorldPop (R2024_v1) 2020 population raster.[3] Population totals were computed for both pre-merger and post-merger ADM3 polygons, enabling direct population-attribution comparison under each configuration. Resulting differences were expressed as percentage changes at the commune level, allowing quantification of how boundary realignments may distort population estimates and affect dependent indicators. Table 2 presents these findings, highlighting communes where administrative restructuring introduces notable population-allocation shifts and potential downstream implications for density calculations, vulnerability assessments and care-demand modelling.

While discrepancies were generally minimal, typically below 5 per cent, in most rural areas, several merged communes in peri-urban Da Nang displayed substantially higher variation, with differences reaching 10–20 per cent (Table 2). These shifts are largely driven by reallocation of densely populated raster cells situated along former boundary edges, where administrative mergers amplify even small geometric adjustments' effects. Despite localized variation, administrative mergers in Da Nang and Quang Nam introduce measurable but manageable changes in spatial population distributions and associated care-service indicators. Through systematic boundary harmonization and transparent methodological documentation, the analysis maintains indicator continuity, minimizes allocation error and ensures comparability with national and international geospatial standards. This approach provides a sound basis for integrating pre- and post-merger data in a consistent manner, supporting reliable interpretation of care demand and supply across the study area.

Table 2: Percentage Differences in Total Population

Commune (Current ADM3)	Population Aggregated from Previous Communes	Population Extracted from Current Communes	Percentage Difference
Xã Sơn Cẩm Hà	12,665	10,612	**+20%** (Increase in aggregated count implies previous boundaries allocated population elsewhere)
Xã Tiên Phước	23,353	25,482	**-8%**
Phường Hòa Khánh	122,875	110,365	**+11%** (Shift from previous boundary aggregation)
Phường Thanh Khê	147,479	163,599	**-10%**

3 At the time of analysis, R2024_v1 was the only version available and was therefore used for comparisons and validation against census data. A newer WorldPop release became available during report finalization, showing minor numerical differences; these do not affect the analysis or conclusions.

III. Mapping the Demand for Care Services

Mapping care-service demand involves identifying geographic areas where care needs are expected to be higher based on population demographic characteristics. Age groups are typically used as proxies for specific care needs. In this analysis, two population groups are prioritized: children of pre-school age (0–5 years), who generally require early childhood care services, and older persons (60 years and above), who may require varying eldercare forms. For standardized indicators, these age ranges may differ, but they provide a robust benchmark for examining care-need spatial patterns in Da Nang.

Demand analysis draws on a range of variables that together offer nuanced understanding of demographic composition, as detailed in **subsection 3.1** Recognizing that care needs do not exist in isolation, the analysis also incorporates contextual factors, such as exposure to climate-related hazards, land-use patterns and living conditions, which are widely documented in global literature as influencing both care-need intensity and households' ability to meet them. These factors are examined in **subsection 3.2** Each subsection is structured around two components: an overview of data needs and sources, followed by data preparation and cleaning procedures. **Subsection 3.3** presents final indicators used for demand analysis, while **subsection 3.4** explains spatial-unit construction.

Both demographic and contextual variables rely exclusively on secondary datasets, detailed in corresponding subsections. Taken together, these components enable multidimensional care-demand analysis that captures not only population characteristics but also the structural and environmental conditions shaping Da Nang's care landscape.

3.1 DEMOGRAPHIC VARIABLES

3.1.1 Identification of data sources

To assess the demand for care services, reliable demographic data are essential. Although the 2019 Population and Housing Census provides the most accurate representation of the population in the area of interest, commune-level data were not available for this analysis. As a result, WorldPop population estimates were used as the primary source for mapping care demand.

While the census offers detailed disaggregated information (e.g., employment status, migration patterns and education levels), which allows a more comprehensive demographic assessment, WorldPop does not include these variables. Nonetheless, it provides high-resolution population estimates that allow for the identification of care-dependent groups at the commune level and are therefore suitable for spatial demand modelling in Da Nang. Table 3 summarizes the data needs and sources used in this study.

Table 3: Data Sources for the Demand for Care Services

Type of data source	Data source	Date (last updated)	Granularity (lowest unit)	Details of the dataset
Census	IPUMS Viet Nam census	2019	District	This dataset includes total population by sex and age.
Administrative data sources	Statistics Office	2021–2024	Commune	This dataset includes total population, as well as population disaggregated by sex.
Census	Statistics Office	2019	City	This dataset includes the population by disability status: disabled; not disabled.
Raster	WorldPop previous stable version	2020	Pixel (~100m)	This dataset is broken down by sex and age groups (0 to 95+, in five-year intervals). It was produced around 2018 as the first generation of global multi-annual gridded population estimates.
Raster	WorldPop R2024B_v1	2020	Pixel (~100m)	This dataset is broken down by sex and age groups (0 to 95+, in five-year intervals). It is a beta release of WorldPop's next-generation global gridded population dataset, covering ~100 m resolution and the years 2015–2030.
Raster	GHS-POP R2023A	2020	Pixel (~100m)	Produced by the European Commission Joint Research Centre under the Global Human Settlement Layer (GHSL) initiative, the GHS-POP R2023A product provides the number of people per cell for years from 1975 through to 2030, at five-year intervals.

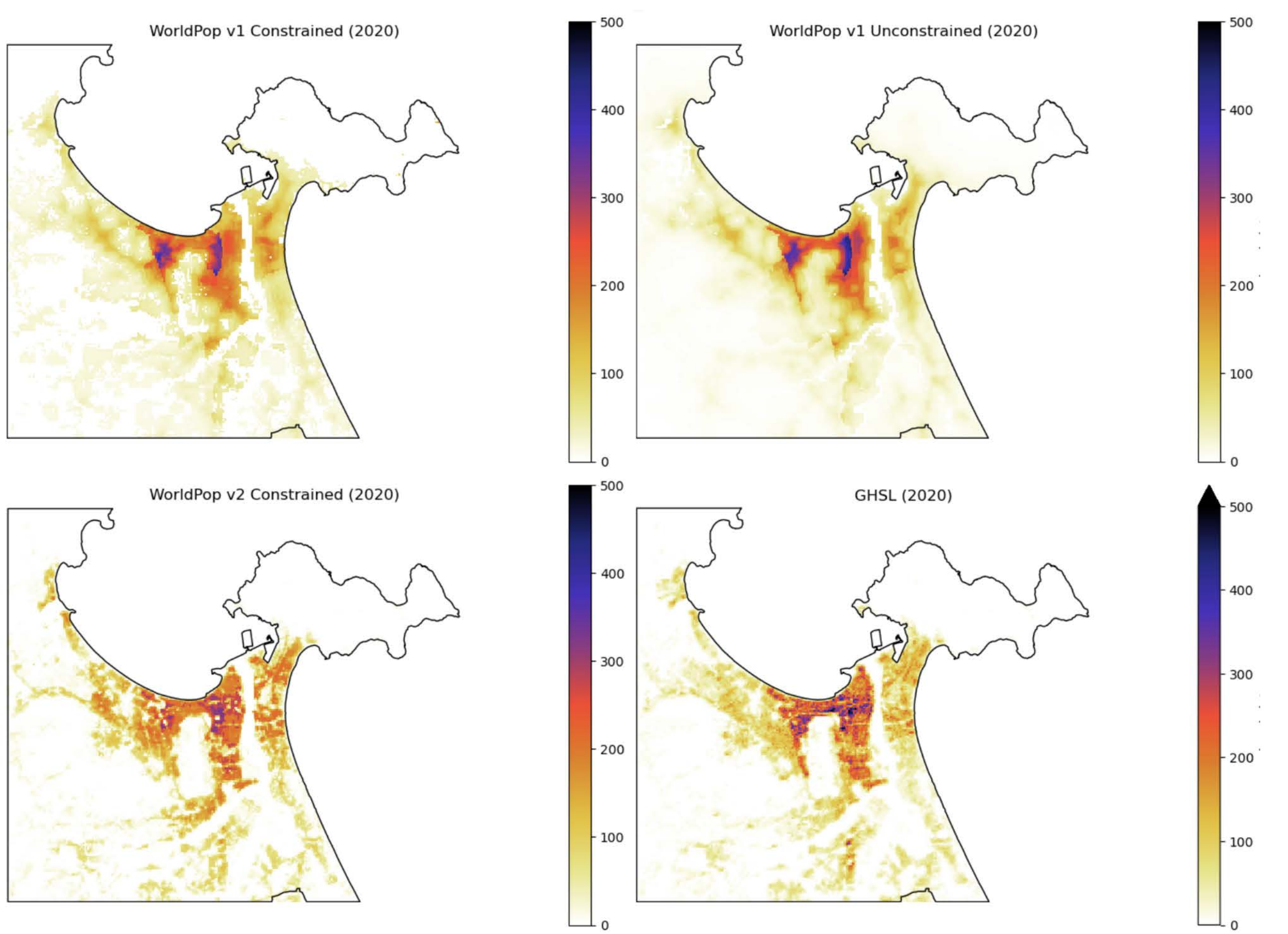
3.1.2 Data sources comparison

This section documents the selection and validation process for the primary gridded population dataset used in this analysis. The dataset needed to provide accurate, high-resolution population estimates with sufficient demographic detail and temporal coverage to support modelling of current and future care service demand. To identify the most suitable source, a comparative assessment was conducted of leading global gridded population products, including the previous stable version of WorldPop, the latest beta release of WorldPop and the Global Human Settlement Layer Population (GHS-POP R2023A). These datasets were evaluated against key criteria such as spatial resolution, methodological transparency, update frequency, internal consistency and their capacity to support disaggregation relevant to gender analysis. This systematic comparison ensured that the selected dataset met the analytical requirements for mapping care demand and aligned with emerging best practices for integrating demographic and contextual layers within GIS-based care system assessments (Table 4).

WorldPop Global 2 (R2024B_v1) was selected as the primary gridded population dataset for this analysis because it incorporates the most recent 2019 census data and provides detailed age- and sex-specific disaggregation essential for modelling care demand. Its spatial granularity is comparable to that of the GHSL Population Layer, as illustrated in Map 3, while offering a level of demographic detail not available in GHSL, which does not include age or sex breakdowns. Earlier versions of WorldPop were excluded because they were constructed using only the 1999 and 2009 censuses and therefore do not reflect current population distributions.

Validation of the selected gridded population product relies on high-quality, locally sourced demographic information, referred to here as "Ground Truth", against which the gridded estimates are systematically compared. Table 5 presents the census and administrative datasets used as both input layers and validation benchmarks for the merged Da Nang City area, ensuring internal consistency and methodological transparency throughout the validation process. The 2020 outputs of the selected gridded population products were validated against two independent reference points: the 2019 Population and Housing Census, used as the baseline benchmark, and the 2021 administrative population totals, which provide an additional consistency check to assess temporal accuracy and alignment with recent demographic trends.

Map 3: Visualization of Gridded Population Products



Source: Authors' elaboration using WorldPop (2018), WorldPop (2024), and GHS-POP (2023)

Table 4: Overview of Ground Truth Demographic Input Data

Attribute	2019 Census Data	Administrative Data
Source	IPUMS	Da Nang City Statistics Office and Quang Nam Provincial Statistics Office
Year(s) Covered	2019	2021, 2022, 2023, 2024
Administrative Level	District (ADM2)*	Pre-merger Commune (ADM3)
Total Population Size near 2020	2019: 2,630,229	2021: 2,713,965

*District-level administration was dissolved following the 2025 city merger.

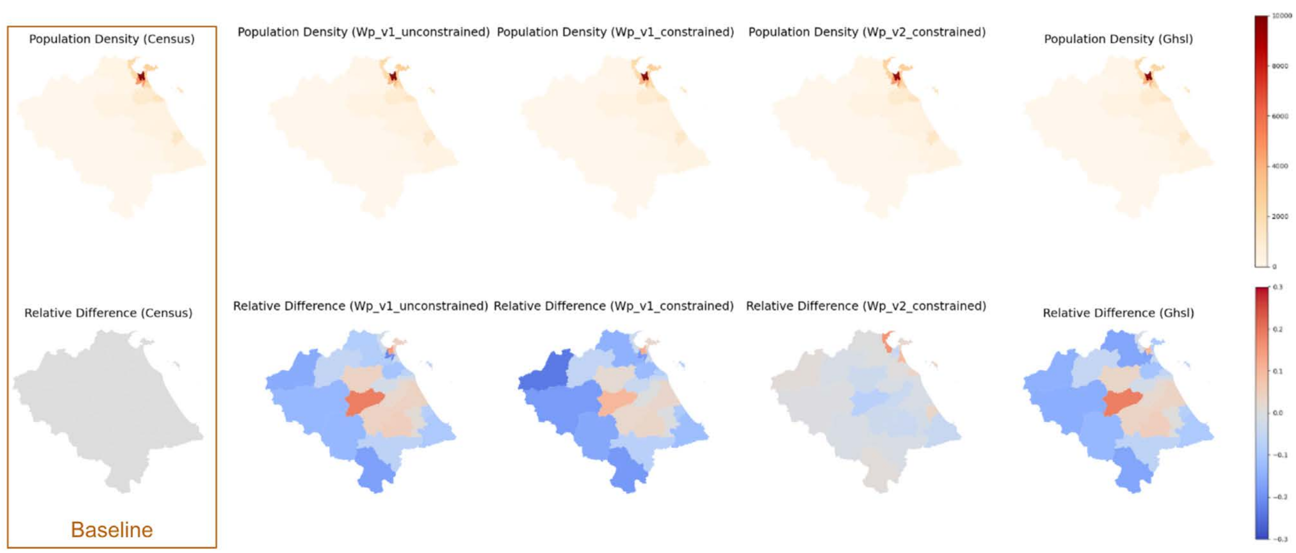
Table 5: Comparative Overview of Gridded Population Products

Feature	WorldPop v1 (2000–2020)	WorldPop R2024B_v1	GHSL (GHS-POP)
Source / Link	WorldPop previous stable version	WorldPop R2024B_v1	GHS-POP R2023A
Spatial Resolution	~100m	~100m	~100m
Temporal Coverage	2000–2020 yearly	2015–2030 yearly	1975–2030 every five years
Input Census Data	1999 and 2009 censuses	2009 and 2019 censuses	2009 census (CIESIN GPWv4.11)
Methodology	Top-down dasymetric mapping using ancillary data (land cover, roads, nightlights); constrained/unconstrained on built settlements	Top-down dasymetric mapping with updated and expanded geospatial covariates, improved approaches to projecting population numbers	Gridded from 1km census counts using built-up area extent from GHSL layers
Age/Sex Structure	**Yes** • Constrained: 2020, 100m • Unconstrained: 2000-2020 yearly, 100m	**Yes** • Constrained: 2015-2030 yearly, 100m	**No**, it is not disaggregated by age/sex
Constrained vs. Unconstrained	**Constrained**: allocation only in built areas **Unconstrained**: distributes based on all input covariates	Only **constrained** version available for urban studies	Similar to **constrained**
Advantages	High resolution; updated annually; useful for retrospective and near-present analysis	Includes future projections; age-sex breakdowns; policy relevance	Long historical coverage; globally consistent; easy to use
Limitations	Older censuses	Same resolution as v1, still at beta test phase	Lacks age/sex; fewer updates; older censuses

Table 6: Comparison of 2020 Grid Estimates vs. Ground Truth (Da Nang and Quang Nam)

	Ground Truth		Gridded Population Products			
Area	2019 Census	2021 Admin Population	2020 WorldPop v1 unconstrained	2020 WorldPop v1 constrained	2020 GHS-POP	2020 WorldPop Global 2 constrained
Da Nang	1,134,332	1,195,488	1,141,228	1,119,063	1,091,673	1,139,157
Quang Nam	1,495,897	1,518,477	1,437,687	1,405,809	1,434,773	1,467,723

Map 4: Population Density and Relative Difference



Source: Authors' elaboration using IPUMS (2019), WorldPop (2018), WorldPop (2024), and GHS-POP (2023)

Spatial consistency was first assessed visually, as illustrated in the first row of Map 4, which shows that WorldPop Global 2 accurately reflects population distribution across the study area. The dataset captures the dense urban concentration characteristic of core Da Nang as well as the more dispersed settlement patterns across Quang Nam. The Relative Difference measure—calculated as (Grid Pop – Census) / Census at the ADM2 level—reveals spatial variations that align with well-documented demographic dynamics in the region. These include the rapid pace of urbanization, with population growth in urban areas exceeding twice the national rate and nearly six times that of rural areas between 2009 and 2019, as well as the pronounced effects of internal migration. Interprovincial migrants represent the largest share of migrants aged five and older, and women constitute more than half (55.5 per cent) of these internal migrants, underscoring the gendered dimensions of population mobility and care demand.

The validation procedures (comparing gridded estimates against both the 2019 Census baseline and the 2021 administrative population totals) further confirm the robustness of the WorldPop Global 2 (R2024B_v1) constrained product. The 2020 output offered the closest numerical fit for both Da Nang (1,139,157) and Quang Nam

(1,467,723), outperforming alternative products such as WorldPop v1 and GHS-POP. Beyond numerical accuracy, the dataset's 100-metre spatial resolution provides the level of granularity required to capture spatial heterogeneity in care-relevant population patterns, including concentrated urban growth corridors and the demographic impacts of interprovincial migration.

3.2 CONTEXTUAL FACTORS

3.2.1 Identification of data sources

To deepen the analysis and more accurately assess care demand, two categories of contextual factors were identified as essential: land use and climate risks (Table 7). These dimensions provide the structural and environmental context shaping how care needs emerge and evolve within Da Nang.

- **Land use:** Analysis of land use enhances the spatial accuracy of care demand assessments by illustrating how population settlement patterns, economic activity and service availability intersect. Land use determines where people live, work and access essential services, making it a critical layer for understanding spatial disparities in care access. Integrating land use data helps distinguish between genuinely underserved areas—where population density and service needs are high but provision is lacking—and areas where limited facilities reflect the nature of land use (e.g., agricultural zones or industrial sites) rather than inequitable distribution. This reduces the risk of misclassification and ensures that policy recommendations remain grounded in the city's territorial logic. By highlighting mismatches between care needs and care service locations, land use analysis supports the development of place-based and equitable care policies that reach the most vulnerable populations.

- **Climate risks:** Climate risks constitute a critical contextual factor shaping care demand in Da Nang City. Situated along Viet Nam's central coastline facing the South China Sea, Da Nang's strategic economic and transport role is coupled with high exposure to climatic and environmental hazards. The city experiences a tropical monsoon climate with distinct wet and dry seasons and is increasingly affected by climate change, which has driven rising temperatures, more intense rainfall events and gradual sea-level rise. These shifts have amplified the frequency and severity of flooding and cyclones, posing significant risks to both urban and peri-urban populations. Heavy rainfall, often triggered or intensified by tropical cyclones, regularly exceeds the capacity of existing drainage infrastructure, resulting in flash floods in low-lying and densely populated areas. Rapid urbanization has compounded these vulnerabilities by reducing natural water absorption and altering hydrological flows. Cyclones striking Viet Nam's central coast frequently affect Da Nang, bringing destructive winds, storm surges and coastal inundation that disrupt transport networks, electricity, water supply and essential public services.

These climate-related hazards directly influence care demand. Older persons, children, persons with disabilities and low-income households face heightened vulnerability during extreme weather events, often requiring additional support during evacuation, sheltering and recovery. Flooding and cyclones can damage or temporarily close childcare, eldercare and healthcare facilities, reducing access to services precisely when needs intensify. Moreover, stagnant water following floods increases the risk of waterborne and vector-borne diseases such as dengue and diarrhoeal illness, placing additional pressure on household caregivers—predominantly women and girls—and on already strained health systems. Integrating climate risk into the analysis therefore provides essential insight into how environmental hazards exacerbate existing inequalities and shape the spatial and temporal distribution of care needs.

Table 7: Data Sources for Contextual Factors

Data need	Data source type	Data source	Date	Granularity (lowest unit)	Details of the dataset
Relative Wealth Index	Raster	Meta Data for Good	2021	Pixel (~2.4km)	This dataset estimates relative living standards across countries by analysing de-identified connectivity data, satellite imagery and other non-traditional data sources.
Land Use	Shapefile	Administrative Data Sources	NA	Polygon	The shapefile includes polygons representing land use areas categorized into different level 1 classes: standing water, flowing water, residential areas, other vegetation, forest, cultivated land, annual crops and perennial crops.
Commercial Activities	Shapefile	Humanitarian OpenStreetMap	2025	Point	The point layer was extracted based on tags indicating commercial use, including shops, markets, restaurants and other business establishments. The extracted points were processed to remove duplicates and ensure spatial consistency, enabling an accurate representation of the distribution and density of commercial activity across the study area.
Flood and Cyclone Risk	Shapefile	National Center for Hydrometeorological Forecasting	2023	Polygon	The shapefile presents spatial variations in multi-hazard risk levels across pre-merger Da Nang and Quang Nam, integrating exposure to tropical cyclones and associated storm surges. Risk levels are classified into five categories: very low, low, medium, high and very high.
NDVI (Normalized Difference Vegetation Index)	Raster	ESA Copernicus Sentinel 2	2024	Pixel (~10m)	The 2024 annual NDVI, derived from Harmonized Sentinel-2 MSI data, measures vegetation conditions. Values range from –1 to 1, where negative values denote water, near-zero values indicate bare surfaces, moderate values (0.2–0.4) represent sparse vegetation and higher values (approaching 1) reflect dense, healthy vegetation.

3.3 DATA PREPARATION AND CLEANING

For effective use of geospatial data, all datasets were standardized to ensure consistency in projection, resolution and spatial alignment. Processing was conducted primarily in Quantum Geographic Information System (QGIS) version 3.40.5, with selected analytical procedures executed in Python to support automation and reproducibility. Following data cleaning and preparation, the finalized datasets were integrated with the corresponding administrative boundary shapefiles, enabling robust spatial analysis across communes and commune clusters.

- **Land use:** Land use analysis was conducted at the administrative commune (ADM3) level. Land use datasets were spatially clipped to commune boundaries to isolate areas falling within each unit. The area of each land use class was then calculated in square kilometres and categorized into eight groups: standing water, flowing water, residential areas, other vegetation, forest, cultivated land, annual crops and perennial crops. This approach provides a detailed spatial profile of land use composition within each commune.
- **Commercial activities:** To characterize the spatial distribution of commercial activity, two complementary indicators were derived from Humanitarian OpenStreetMap (OSM) data: *commercial facilities per capita and commercial facilities per km²*. All OSM commercial activity points were first extracted and aggregated by administrative unit. For the per capita measure, the total number of commercial facilities was

divided by the population of each unit. For the density indicator, the same facility count was divided by the area of the administrative unit in square kilometres.

- **Flood and cyclone combined risk:** Combined flood and cyclone risk was assessed using spatial data on tropical cyclone exposure and associated storm surge zones for pre-merger Da Nang and Quang Nam. Raw risk maps were digitized and merged, after which the average risk level was calculated for each post-merger commune. Risk levels were harmonized into five categories: very low, low, medium, high and very high, reflecting the relative exposure of each commune to hydrometeorological hazards.

- **Population exposed to flood hazards:** Population exposure to flood and cyclone risk was calculated at the cluster level to align vulnerability metrics with care demand indicators. Only communes classified as high or very high risk were included, enabling a focused examination of the most affected areas. For each cluster, the total population residing in these high-risk communes was summed and divided by the cluster's total population to derive the proportion of people exposed to elevated risk.

3.4 FINAL INDICATORS

Table 8 presents the demographic and contextual indicators used in this study, outlining for each indicator its definition, calculation method and corresponding data sources, thereby providing a transparent reference for how the variables were constructed and operationalized within the analysis.

Table 8: Final Indicators for the Analysis of Demands for Care Services

Category	Indicator	Description	Data Source
Demographic Demographic	Total Population (2020, constrained)	Estimated total population (constrained) from WorldPop 2020 dataset.	WorldPop R2024B_v1
	Population Density (2020, constrained)	Population per square kilometre (2020, constrained).	WorldPop R2024B_v1
	Female Population (2020, constrained)	Estimated female population (constrained) from WorldPop 2020.	WorldPop R2024B_v1
	Male Population (2020, constrained)	Estimated male population (constrained) from WorldPop 2020.	WorldPop R2024B_v1
	Children (0–4 years)	Population aged 0–4 years.	WorldPop R2024B_v1
	Female Children (0–4 years)	Female population aged 0–4 years.	WorldPop R2024B_v1
	Male Children (0–4 years)	Male population aged 0–4 years.	WorldPop R2024B_v1
	Child Dependents (0–14 years)	Population aged 0–14 years.	WorldPop R2024B_v1
	Female Child Dependents (0–14 years)	Female population aged 0–14.	WorldPop R2024B_v1
	Male Child Dependents (0–14 years)	Male population aged 0–14.	WorldPop R2024B_v1
	Elderly Population (60+)	Population aged 60 years and above.	WorldPop R2024B_v1
	Female Elderly (60+)	Female population aged 60 years and above.	WorldPop R2024B_v1
	Male Elderly (60+)	Male population aged 60 years and above.	WorldPop R2024B_v1
	Elderly Dependents (65+)	Population aged 65 years and above (dependents).	WorldPop R2024B_v1
	Female Elderly Dependents (65+)	Female population aged 65 years and above.	WorldPop R2024B_v1
	Male Elderly Dependents (65+)	Male population aged 65 years and above.	WorldPop R2024B_v1
	Working-Age Population (15–64)	Population aged 15–64.	WorldPop R2024B_v1
	Female Working-Age Population (15–64)	Female population aged 15–64.	WorldPop R2024B_v1
	Male Working-Age Population (15–64)	Male population aged 15–64.	WorldPop R2024B_v1

Category	Indicator	Description	Data Source
Demographic	Women of Reproductive Age (15–49)	Female population aged 15–49 years.	WorldPop R2024B_v1
	Total Population (2025, admin website)	Total population for 2025 (from administrative website).	WorldPop R2024B_v1
	Administrative Area (2025, admin website)	Area of the administrative unit (km²) based on administrative website.	WorldPop R2024B_v1
	Population Density (2025, admin website)	Population per km² for 2025 (from administrative website).	WorldPop R2024B_v1
	Total Dependents	Sum of child and elderly dependents (0–14 + 65+).	WorldPop R2024B_v1
	Child Dependency Ratio	Ratio of children (0–14) to working-age population (15–64).	WorldPop R2024B_v1
	Elderly Dependency Ratio	Ratio of elderly (65+) to working-age population (15–64).	WorldPop R2024B_v1
	Total Dependency Ratio	Ratio of total dependents (0–14 & 65+) to working-age population.	WorldPop R2024B_v1
	Child-Woman Ratio	Number of children (0–4) per 1,000 women aged 15–49.	WorldPop R2024B_v1
Contextual - Socio-economic	Relative Wealth Index (Meta)	Mean relative wealth index (Meta model).	Meta Data for Good
Contextual - Land Use	Water Share	Share of water area within the administrative boundary.	Administrative Data Sources
	Agricultural Share	Share of agricultural area within the administrative boundary.	Administrative Data Sources
	Forest Share	Share of forest area within the administrative boundary.	Administrative Data Sources
	Residential Area Share	Share of residential land within the administrative unit.	Administrative Data Sources
	Other Vegetation Share	Share of non-forest vegetation within the administrative unit.	Administrative Data Sources
	Other Land Share	Share of other land types (e.g., barren, mixed).	WorldPop R2024B_v1
	Commercial Facilities per Capita	Number of commercial points per person.	Humanitarian OpenStreetMap
	Commercial Facilities per km²	Number of commercial facilities per square kilometre.	Humanitarian OpenStreetMap
Contextual - Climate	NDVI (All Areas)	Mean NDVI index across all areas.	ESA Copernicus Sentinel 2
	NDVI (Urban Areas)	Mean NDVI index (urban areas only).	ESA Copernicus Sentinel 2
	Flood and Cyclone Risk Level	Average hazard index combining flood and cyclone risk.	Administrative Data Sources, WorldPop R2024B_v1
	Total Population (2020, constrained) under High Flood/Cyclone Risk (%)	Share of total constrained population (2020) living in high flood/cyclone risk areas.	Administrative Data Sources, WorldPop R2024B_v1
	Female Population (2020, constrained) under High Flood/Cyclone Risk (%)	Percentage of total constrained female population (2020) in high-risk areas.	Administrative Data Sources, WorldPop R2024B_v1
	Male Population (2020, constrained) under High Flood/Cyclone Risk (%)	Percentage of total constrained male population (2020) living in high-risk areas.	Administrative Data Sources, WorldPop R2024B_v1
	Children (0–4) under High Flood/ Cyclone Risk (%)	Percentage of total population aged 0–4 exposed to flood/cyclone risk.	Administrative Data Sources, WorldPop R2024B_v1
	Female Children (0–4) under High Flood/Cyclone Risk (%)	Percentage of female children aged 0–4 exposed to flood/cyclone risk.	Administrative Data Sources, WorldPop R2024B_v1
	Male Children (0–4) under High Flood/ Cyclone Risk (%)	Percentage of male children aged 0–4 exposed to flood/ cyclone risk.	Administrative Data Sources, WorldPop R2024B_v1
	Elderly Population (60+) under High Flood/Cyclone Risk (%)	Percentage of total population aged 60+ exposed to high climate risk.	Administrative Data Sources, WorldPop R2024B_v1
	Female Elderly (60+) under High Flood/ Cyclone Risk (%)	Percentage of female elderly (60+) exposed to flood/ cyclone risk.	Administrative Data Sources, WorldPop R2024B_v1
	Male Elderly (60+) under High Flood/ Cyclone Risk (%)	Percentage of elderly male population (60+) exposed to flood/cyclone risk.	Administrative Data Sources, WorldPop R2024B_v1
	Percentage of Childcare Facilities under High Flood/Cyclone Risk (%)	Percentage of childcare facilities exposed to flood/ cyclone risk.	Administrative Data Sources, WorldPop R2024B_v1
	Percentage of Old-age care Facilities under High Flood/Cyclone Risk (%)	Percentage of old-age care facilities exposed to flood/ cyclone risk.	Administrative Data Sources, Google Places API, Humanitarian OpenStreetMap

Source: Elaborated by the authors

IV. Mapping the Supply of Care Services

Mapping the supply of care services involves identifying the geographic distribution of services available for at least two key population groups: children of pre-school age and older persons. This subsection outlines the methodological steps required to develop a spatial representation of care service availability across the study area. The process consists of four sequential components: identifying data needs and corresponding data sources; preparing and standardizing datasets for spatial analysis; constructing the final indicators used to measure care service provision; and integrating these indicators into the broader geospatial framework. Together, these steps provide a transparent and replicable basis for assessing the adequacy, accessibility and spatial equity of care services across communes and clusters.

4.1 IDENTIFICATION OF DATA SOURCES

To develop a reliable spatial representation of care supply, the analysis draws on three primary categories of data: Google Places API data, crowdsourced datasets and administrative geo-referenced data (Table 9).

- **Google Places API data** play a central role in identifying care facilities where no official geo-referenced datasets exist. This is particularly relevant for eldercare services, for which preliminary desk research identified no comprehensive secondary data sources. In such cases, information was obtained through web scraping of Google Maps, enabling the extraction of facility names, types and coordinates that would otherwise be inaccessible. This method ensures broader coverage of services, especially in contexts where institutional registries are incomplete or unavailable.
- **Crowdsourced data, primarily from OpenStreetMap (OSM)**, provide an additional openly accessible source of geospatial information. OSM datasets include the locations and attributes of educational institutions, such as schools and early childhood facilities, as well as health-related points of interest, including hospitals, clinics and pharmacies. OSM also captures amenities relevant to older persons' well-being, such as recreational facilities and retirement homes. While valuable for its breadth and frequent updates, the completeness and accuracy of OSM data vary depending on local contributor engagement and validation efforts, requiring careful cross-checking against other data sources.
- **Administrative data** provide the most formal record of care facilities and were obtained through the Da Nang Women's Union, drawing on information compiled from the Department of Education and Health and other relevant government and institutional sources.

These datasets offer structured insights into service provision but vary in completeness across categories, with gaps in geographic coordinates, facility capacity and staffing information. Moreover, administrative datasets were available only for the pre-merger Da Nang area, limiting direct comparability across the broader study region.

Together, these three data sources provide a complementary foundation for constructing a comprehensive and spatially coherent picture of care service availability. Their integration strengthens the reliability of the final indicators used to assess care supply across communes and clusters.

Table 9: Data Sources for Mapping the Supply of Care Services

	Type	Data source	Unit	Date	Categories
Google Places API (Web scraping)					
Child care services	Web scraping	Google Places API	Point data	2025	No categories, classified by name
Old-age care services	Web scraping	Google Places API	Point data	2025	No categories, classified by name
PwD care services	Web scraping	Google Places API	Point data	2025	No categories, classified by name
Crowdsourced data					
Education facilities	Crowd sourced, Geo-located data	OpenStreetMap Export (HDX)	Point data	2024	1.'Kindergartens' 2.'Schools'
Health facilities	Crowd sourced, Geo-located data	OpenStreetMap Export (HDX)	Point data	2024	1.'Social Facility' 2.'Nursing Home'
Points of interest	Crowd sourced, Geo-located data	OpenStreetMap Export (HDX)	Point data	2024	1.'Childcare' 2.'Nursing Home' 3.'Social Facility' 4.'Community Centre' 5'Social Centre'
Administrative data					
Care facilities	Administrative data	Department of Education and Health	Point data (no geolocation)	2024	Available for Da Nang. 1. Education and health centres. 2. Care facilities 3. Social assistance facilities

4.2 DATA PREPARATION AND CLEANING

4.2.1 Google Places API

Step 1 Extraction: The Google Places API provides several search functionalities, including "nearby search" and "text search". For this study, the nearby search function was selected because it enables identification of places of interest within a specified radius from a given location. Using this method, the API returns establishments whose names, categories or metadata match a predefined set of keywords (Table 10). This capability is essential in contexts where no consolidated geo-referenced data sources exist for childcare or eldercare facilities.

The keyword-based nature of the API introduces several limitations. The underlying metadata—comprising elements such as name, location, establishment type and user-generated descriptions—may be incomplete or inconsistently tagged. As a result, variations in how facilities are labelled can lead to irrelevant matches or, conversely, the exclusion of relevant facilities not explicitly associated with the selected keywords. This reliance on metadata completeness and consistency may introduce bias, necessitating careful validation and cleaning.

A further limitation is the platform's cap of 60 returned places per query. When a query radius is too large, the API may truncate results, omitting facilities that exceed this threshold. To mitigate this, a systematic grid-based approach was adopted. A rectangular grid spaced at 1.5 kilometres was generated to cover the geographical area under analysis, producing a total of 280 reference points. Although some points fall outside administrative boundaries, they were included to maintain grid regularity. A nearby search with a 1-kilometre radius was performed around each point. Combining 1.5-kilometre grid spacing with a 1-kilometre search radius ensures full geographic coverage with intentional overlap, reducing the risk of omission. Overlapping results produce duplicates, which are addressed during data cleaning.

To further counteract the 60-result cap, all locations where the API returned exactly 60 facilities were re-queried using a finer local grid. This refinement captures additional establishments not retrieved in the initial pass, ensuring that high-density service areas are comprehensively mapped. The final dataset emerging from this multi-stage process contains a substantial number of duplicates due to overlapping search areas, making cleaning and deduplication a critical step in ensuring an accurate and reliable inventory of care-related facilities.

Step 2 Deduplication: Deduplication followed a structured three-criteria process designed to ensure each establishment is represented once while preserving distinct facilities.

- The **first criterion** relies on the *place_id* provided by the Google Places API, which serves as a unique identifier for each establishment; only a single record per *place_id* is retained.
- The **second criterion** uses geospatial proximity to detect potential duplicates. Facilities are compared using latitude and longitude values across two precision thresholds. The first rounds coordinates to six decimal places (approximately 11.1 centimetres), capturing extremely close matches that almost certainly represent duplicates. The second rounds to five decimal places (approximately 1.11 metres), allowing identification of slightly displaced coordinates that may refer to the same establishment. A custom script identifies clusters of facilities within these thresholds, after which human judgment is applied to compare facility names. Where names are highly similar, additional verification is conducted using Google search results to determine whether listings correspond to one or multiple distinct facilities. Where names

differ and verification confirms distinct entities, both records are retained.

- The **third criterion** addresses the accuracy of facility classifications. Each facility includes a type variable assigned through Google's metadata. Some facilities relevant to childcare or eldercare appear under unexpected or 'suspicious' types, such as a childcare centre listed as a 'clothing store', due to metadata inconsistencies. Facilities with such types are isolated for manual review. Verification relies first on the facility name; if the name does not clearly indicate a relationship to care services, a supplementary Google search is conducted to confirm the establishment's function. Only facilities that demonstrably meet the criteria for childcare or eldercare services are retained in the final dataset.
- This multi-step deduplication process ensures a high-quality, accurate and reliable dataset for mapping care service availability, minimizing false inclusions and omissions while maintaining methodological transparency.

Step 3 Filtering: Because of the keyword-based nature of Google Places API searches, the raw dataset often contains establishments that are not relevant to childcare or eldercare services. For example, an airport with a 'children's area' may include terms associated with childcare in its metadata and thus be returned as a match, despite not providing care services. A systematic filtering process is therefore required to remove such irrelevant entries. The initial cleaning stage consists of two stages.

- The **first stage** excludes places whose names or assigned types clearly indicate that they are not care-related. Categories such as libraries, tourist attractions, restaurants, law offices, movie theatres, shopping malls, airports and supermarkets are removed at this stage. However, some irrelevant establishments remain because non-care facilities may contain ambiguous or misleading terms in their metadata. These are identified through closer inspection of facility names or manual verification using Google search results.
- The **second stage** applies a keyword-based filter to remove places whose names do not include relevant terms linked to childcare or eldercare, such as 'school', 'kindergarten' or 'nursery'. While effective in narrowing the dataset, this approach risks excluding legitimate establishments whose names do not explicitly reference care, including facilities named after individuals or geographic locations. For this reason, filtering criteria are refined and supported by targeted manual review. In many cases, confirming the nature of a facility requires an individual online search to verify whether it provides services relevant to this study.

This combination of automated and manual filtering ensures that the final dataset includes only genuine care-related establishments while minimizing the risk of excluding legitimate facilities.

Table 10: List of Keywords

English	Vietnamese Type
Childcare	
Childcare	Trung tâm chăm sóc trẻ
Kindergarten	Trường mẫu giáo
Preschool	Trường mầm non
Childcare centre	Trung tâm giữ trẻ
Nursery	Nhà trẻ
Daycare centre	Trung tâm giữ trẻ
Special education school	Trường giáo dục đặc biệt
Childcare group	Nhóm chăm sóc trẻ em
Home-based child care group	Nhóm trông trẻ tại nhà
Private kindergarten	Trường mầm non tư thục
PwD	
Social protection facility	Cơ sở bảo trợ xã hội
Support service facility for persons with disabilities	Cơ sở dịch vụ hỗ trợ người khuyết tật
Care centre/facility for persons with disabilities	Trung tâm/Cơ sở chăm sóc người khuyết tật
Rehabilitation centre for persons with disabilities	Trung tâm phục hồi chức năng người khuyết tật
Mental health care facility	Cơ sở chăm sóc người tâm thần
Special education centre for children	Trung tâm giáo dục trẻ đặc biệt
Old age care	
Elderly care facility	Cơ sở chăm sóc người cao tuổi/người già
Elderly Care	Cơ sở chăm sóc người cao tuổi/người già
Home-based healthcare services for the elderly	Dịch vụ chăm sóc sức khỏe tại nhà cho người cao tuổi
Elderly nursing centre	Trung tâm điều dưỡng người cao tuổi
Retirement home	Nhà hưu trí
Nursing home	Nhà dưỡng lão
Elderly care centre	Trung tâm chăm sóc người cao tuổi
Daycare centre for elderly	Trung tâm chăm sóc ban ngày
Health centre (integrated)	Trung tâm y tế

4.2.2 Crowdsourced data

This study draws on three crowdsourced datasets from Humanitarian OpenStreetMap (OSM): education facilities, health facilities and points of interest. Each dataset includes an “amenity” field that specifies the type of facility. As an initial step, all datasets were spatially filtered to retain only facilities located within the study area.

Step 1 Extraction:

- For the **education dataset**, the analysis focuses on facilities providing childcare or early education. Entries categorized as ‘university’ were excluded because they fall outside the scope of the care supply analysis. Facilities listed as ‘kindergarten’ (5 entries) and ‘school’ (52 entries) were retained, as these may include pre-school or early childhood education services relevant to the target age group.

- For **health facilities**, the raw dataset includes 7 entries classified as 'dentist', 7 as 'doctors', 24 as 'hospital', 52 as 'pharmacy', 4 as 'clinic' and 36 entries without an assigned amenity value. From this set, only 'clinic', 'doctors' and 'hospital' categories were retained because these facility types provide healthcare services relevant to older persons and caregiving needs. Records with missing amenity values were retained for closer inspection during later validation steps.
- The points of interest dataset contains **several amenity types** that complement those identified in the education and health datasets. To broaden the scope of the care supply analysis, amenities categorized as 'childcare', 'prep_school', 'nursing_home', 'community_centre', 'social_center' and 'social_facility' were included because of their relevance to community support services and potential caregiving functions.

Step 2 Deduplication: A structured deduplication process was applied to enhance dataset integrity. Each facility in the OSM-derived datasets contains a unique identifier in the place_id field, which served as the basis for removing duplicate entries. As this field contained no missing values, deduplication consisted of identifying and eliminating repeated records with the same place_id. This procedure was applied consistently across all facility categories, including childcare centres, schools, old age care facilities and facilities serving people with disabilities. The result is a cleaned and consolidated dataset suitable for subsequent spatial analysis of care service availability.

4.2.3 Da Nang's Administrative Data

Step 1 Geocoding: Care facility administrative data were first consolidated into a standardized table to ensure consistency across fields, including institution name, address, district and source table name (in Vietnamese). Geographic coordinates (latitude and longitude) were then assigned through automated geocoding using the Google Maps Geocoding API, enabling efficient spatial referencing of facilities with minimal manual intervention. Of the 906 facilities included in this dataset, 39 (4.3 per cent) could not be successfully geocoded due to incomplete or ambiguous address information. These entries were flagged for potential manual review but excluded from subsequent spatial analysis.

Step 2 Categorization: Each establishment was then assigned to a standardized care supply classification (e.g., *childcare, elderly_care, pwd_care*), while retaining the original Vietnamese label in a separate field to preserve traceability and support future verification. Categorization relied on a keyword-based review of source reference table names. Childcare facilities containing the term 'Trường mầm non' were classified as school-based (261 facilities), while those labelled 'Nhóm/lớp độc lập' or 'Cơ sở giáo dục mầm non độc lập' were categorized as non-school-based (631 facilities). Two facilities listed under 'Hội Bảo trợ người khuyết tật và trẻ mồ' were identified as childcare providers specifically serving children with disabilities and were therefore assigned to the disability-related care category. Ownership information (public/private) was available for fewer than half of the facilities and was not used as a classification variable.

A similar keyword-based approach was applied to identify eldercare facilities within the administrative dataset. Six home-care providers were identified using the keyword 'Giúp việc gia đình', three nursing homes using 'Chăm sóc người cao tuổi', one facility offering combined disability and eldercare services and two establishments classified under general care categories ('Cơ sở chăm sóc khác'). These classifications support the construction of care supply indicators that are consistent, comparable and suitable for integration into the broader geospatial analysis.

4.3. DATA INTEGRATION

To produce a comprehensive master dataset for each target population, namely children, older persons and persons with disabilities, facility information from administrative, Google Places and OSM sources was systematically combined into unified databases. Integration followed a structured three-stage procedure designed to ensure consistency across sources, remove duplicates and retain as much spatial detail as possible.

- **Step 1: Preprocessing** To enable comparability across datasets, all facility categories extracted from Google Places and OSM were first harmonized with the administrative classification framework. For childcare services, facilities labelled as "kindergarten", "preschool" and "montessori" were recoded as school-based childcare, while those listed as "nursery" or "community center" were classified as non-school-based. Facilities tagged as "special school", "child malnutrition center" or "special care center" were mapped to child disability care. For eldercare services, categories such as "retirement house", "nursing home" and "home care" were retained without modification, as they directly aligned with the study's conceptualization of eldercare supply. Since Google and OSM do not provide a distinct category for disability-related eldercare, additional disability-oriented facilities (e.g., "center for mental disabilities") were grouped under other care services for persons with disabilities.

- **Step 2: Integration of Google API and Crowdsourced Data** Given the small number of relevant OSM records, each OSM facility was individually inspected in QGIS and cross-referenced against establishments identified through the Google API. Matches were confirmed based on name similarity and spatial proximity, after which duplicates were consolidated into a unified web-sourced dataset. This integrated dataset includes 1,070 childcare facilities, 4 eldercare facilities and 2 facilities serving persons with disabilities identified through web scraping.

- **Step 3: Integration of Administrative Data** Administrative records were incorporated through a two-stage deduplication process designed to address name inconsistencies and variations in coordinate accuracy. The first method used spatial proximity combined with moderate name similarity: facilities exhibiting at least 50 per cent similarity in Vietnamese names and located within 500 metres of one another were flagged. Manual verification confirmed 109 duplicates. The second method applied stricter name similarity (60 per cent or greater), paired with district-level alignment, leveraging the greater reliability of administrative district information despite known geocoding inaccuracies. This method identified an additional 29 duplicates.

All confirmed duplicates were removed from the Google–OSM dataset to avoid inflating the number of facilities. For the 264 childcare facilities lacking valid geographic coordinates, proportional allocation was used to distribute these facilities across their respective ADM3 units, allowing their inclusion in aggregate analyses. However, the absence of coordinates prevents these facilities from appearing in point-based spatial visualizations. These non-geocoded records represent 14.5 per cent of all childcare facilities. Table 11 summarizes the data sources, total number of records and final counts of care-related facilities.

Table 11: Overview of Data Sources and Care Facilities Identified, by Data Sources

Data Sources	Care Facilities Available Before de-duplication	Data source Total Care Facilities After de-duplication
Childcare		
Administrative data	894	1,826
Google Places API	1,158	
OSM	6	
Eldercare		
Administrative data	12	16
Google Places API	3	
OSM	1	
Care for people with disabilities*		
Administrative data	1	2
OSM	1	

* Note: 'Care for people with disabilities' refers to general disability-care facilities. Age-specific facilities (e.g., child disability centres, disability services for older persons) are combined with the corresponding childcare and eldercare categories, as the analysis integrates disability-specific services with the broader age-based care facility counts.

V. Accessibility

This study assesses gaps between the demand for and supply of care services in Da Nang using AccessMod, a tool developed by the World Health Organization and widely applied to analyse spatial accessibility. Although originally created for the health sector, AccessMod is increasingly used to assess access to a broad range of essential services, making it well suited for understanding care accessibility in rapidly urbanizing contexts such as Da Nang.

The model incorporates key geographic and infrastructural factors, including road networks, topography, physical barriers, population distribution and the location of care facilities, to estimate how easily residents can reach essential services. Two indicators guide the analysis:

- **Travel time**, which measures how long it takes residents to reach the nearest care provider; and
- **Uncovered population**, referring to the share of residents who have no care facility within an accepted travel time threshold.

Consistent with national mobility patterns, motorcycles remain the predominant mode of travel in Viet Nam. They therefore form the basis for travel-time modelling across road categories. Empirical evidence from Hanoi indicates an average motorcycle speed of approximately 20 km per hour, while studies in Da Nang show road speeds ranging from 20–26 km per hour depending on traffic flow and infrastructure quality. Reflecting this evidence, AccessMod assigns road travel speeds of 35 km per hour on primary roads, 20 km per hour on secondary roads, and 15 km per hour on tertiary roads.

Across all non-road land covers, movement is modelled using walking as the primary mode of travel. This reflects Viet Nam–based studies showing that older adults often walk at around 3 km per hour. To anchor the model in inclusive accessibility assumptions, a walking speed of 3 km per hour is applied for both older persons and children below age six. This conservative assumption aligns with national evidence on slower movement among older adults and ensures that travel-time modelling reflects the mobility needs of groups that often require supported travel.

Based on regional evidence and local studies on care access, a maximum travel time of 15 minutes is applied for both childcare and eldercare facilities. These thresholds reflect realistic expectations for accessing essential services in Da Nang, where motorcycles enable relatively efficient intra-city travel but geographic disparities in service provision persist.

Finally, the study examines the spatial alignment between population groups and the distribution of care facilities. Due to data limitations, facility capacity, such as staffing levels or service loads, could not be incorporated. The analysis therefore focuses on geographic accessibility, identifying areas where women, older persons and caregivers may face persistent barriers in reaching essential care services.

Table 12: Data Inputs for AccessMod

Indicator	Type	Data source	Date	Resolution	Use
Demand and Supply					
Child population	Raster	WorldPop	2020	100m	
Population in old age	Raster	WorldPop	2020	100m	
Childcare facilities	Vector	Merged datasets	Multiple	Facility location	
Eldercare facilities	Vector	Merged datasets	Multiple	Facility location	
Other inputs					
Digital Elevation Model (DEM)	Raster	CGIAR-CSI	2018	90m	Used to account for terrain slope, allowing travel speeds to vary depending on the steepness of the land.
Road network	Vector	Humanitarian OpenStreetMap (HOT)	2025	Roads	Used to expand the catchment area of a facility where travel is relatively quicker, reflecting faster movement along certain routes.
Land cover distribution grid /	Raster	JAXA Land Use	2023	10m	Used to describe key features of the study area that are likely to influence travel time for individuals moving through it.
Barriers to movement	Raster	JAXA Land Use	2023	10m	Used to identify areas that are inaccessible to the population, including water bodies and other impassable land covers.
Zone boundaries	Vector	Administrative Commune Boundaries	2024	Communes	

VI. Limitations

6.1 GENERAL LIMITATIONS

Lack of old-age facility data and data on persons with disabilities: A major limitation of this study is the absence of a comprehensive, systematically structured dataset detailing facilities that specifically serve older persons and persons with disabilities. Available administrative and crowdsourced data sources tend to focus on general healthcare or broad social service categories within points of interest datasets, without sufficiently granular classification for these population groups. In the absence of official facility lists, a keyword-based identification approach was employed to locate relevant services. While this method allows for partial reconstruction of the care landscape, it is inherently imprecise. Facilities may be misclassified or entirely omitted due to inconsistent naming conventions, incomplete metadata or gaps in publicly available information. As a result, services for older persons and persons with disabilities are likely underrepresented in the final dataset, affecting the completeness of supply-side indicators.

Lack of location data from government sources: Accurate spatial accessibility modelling depends on the availability of precise geographic coordinates for care facilities. However, inconsistencies and gaps in geospatial information significantly constrained the analysis. Missing or imprecise location data limited the application of WHO AccessMod, an essential tool for modelling travel time and service accessibility, thereby reducing the analytical depth that could be achieved. To mitigate this limitation, alternative data sources such as OpenStreetMap and Google Places API queries were used to approximate facility locations. While these platforms helped address critical data gaps, they may not capture all relevant facilities and can introduce spatial inaccuracies due to uneven metadata quality or contributor activity. These limitations may result in incomplete spatial coverage and reduced reliability of accessibility-related findings.

6.2 LIMITATIONS RELATED TO MEASURING THE DEMAND SIDE

This study faced three key challenges when measuring the demand for care services:

1. **Challenges in Estimating Population Distribution:** Accurate population distribution is fundamental for assessing accessibility and modelling care demand. WorldPop was selected for this study because its estimates align closely with Viet Nam's 2019 census figures and provide the necessary spatial resolution for commune-level analysis. Nevertheless, WorldPop relies on modelling techniques rather than direct enumeration, meaning some degree of inaccuracy is unavoidable. In addition, the dataset does not include certain demographic variables, such as migration status, employment or detailed disability information. This limits the granularity of the care demand assessment and necessitates the use of supplementary data sources to approximate these characteristics.

2. **Wealth Distribution:** Income and wealth data are critical for understanding care needs,

particularly because economic inequality shapes access to services, exposure to vulnerabilities and the capacity of households, especially women, to meet care responsibilities. However, the wealth datasets available for this analysis provide estimates only at the national level, resulting in limited intra-city variability. This constrains the ability to capture neighbourhood-level socioeconomic disparities that significantly influence care demand and may obscure localized pockets of vulnerability in rapidly urbanizing areas.

3. **Data Gaps and Resolution Limitations in Climate Risk Data:** The assessment of climate-related risks, including flood exposure, heat stress and air pollution, is constrained by both data gaps and spatial resolution limitations. Available Urban Heat Island indices do not adequately capture microclimate conditions in smaller or more rural contexts, requiring reliance on vegetation-based proxies that provide only partial insight into spatial temperature variation. Air quality analysis faces similar constraints: the PM2.5 dataset is produced at a coarse spatial resolution, limiting the ability to examine pollution exposure at neighbourhood level, particularly in dense or lower-income urban areas where care needs may be elevated. These constraints affect the precision with which climate-related risks can be integrated into care demand modelling.

6.3 LIMITATIONS RELATED TO MEASURING THE SUPPLY SIDE

A major challenge in facility identification is inconsistency observed across data sources. These discrepancies result in partial and uneven coverage, particularly across key study cities, where some facility types are well documented while others are sparsely represented. Such inconsistencies limit the ability to construct a uniformly complete and comparable care facility database across the full study area.

The diversity of administrative datasets introduces substantial challenges during data cleaning and integration. These datasets are often presented in Vietnamese and vary widely in structure, terminology and level of detail. Standardizing them into a common analytical format requires extensive processing. Overlaps across sources further complicate cleaning, particularly in the childcare sector, where the same facility may appear in multiple registries under slightly different names or classifications. Merging records based on facility names or coordinates is also problematic: automated translations of Vietnamese names may introduce bias, while many datasets lack key metadata such as ownership status, capacity or complete address information.

Variation in categorization practices across administrative sources and cities adds another layer of complexity. Terms such as “community” and “public” are frequently used interchangeably, creating ambiguity in interpreting facility ownership and management structures. In the private early childhood development sector, distinctions between institutional and non-registered facilities are inconsistently applied, leading to misalignment in facility classification across locations. These inconsistencies affect comparability across communes and reduce the precision of supply-side indicators.

Google Places API data introduces additional challenges related to misclassification. Because the API incorporates user-generated content, including reviews, irrelevant facilities, such as parking areas or commercial amenities associated with larger establishments, may be incorrectly identified as care-related services. These inaccuracies clutter the dataset and obscure meaningful spatial patterns, requiring extensive manual validation. The metadata returned by the API often lacks sufficient detail to distinguish between service types and does not consistently differentiate between public and private facilities or provide robust indicators of service type or capacity. This lack of granularity limits the ability to categorize facilities accurately and requires substantial manual filtering and verification to ensure that only correctly classified establishments are retained. These limitations underscore the need to complement Google API data with more structured data sources whenever possible.

Google Places data is also subject to several forms of bias. Because it relies heavily on user-generated inputs, listings may be incomplete, outdated or unevenly distributed across neighbourhoods. Facilities with greater digital visibility, often located in higher-income or more commercial areas, are more likely to appear, while those in lower-income or peri-urban zones may be underrepresented. This bias can distort the spatial representation of care services across the study area and potentially skew analyses of accessibility and service coverage.