Photo ©UN Women/Rawyan Shayema

TOWARDS CARING CITIES: A GEOSPATIAL ANALYSIS IN DHAKA CITY

Methodological Note

DATA-POP ALLIANCE

UN WOMEN

I. Introduction

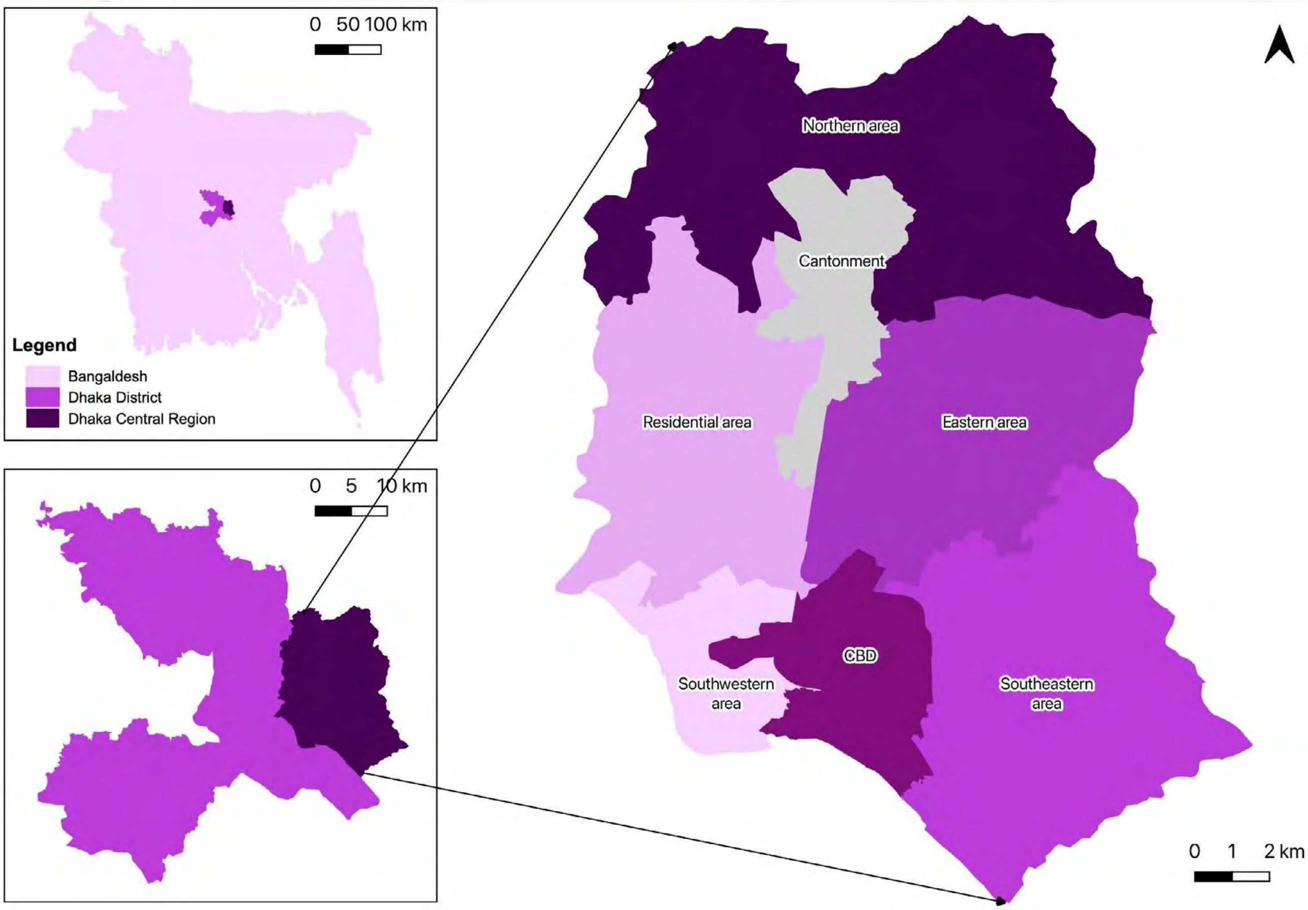
This methodological note explains the technical approach to the Geospatial Analysis of Care Systems in Bangladesh prepared by Data-Pop Alliance with UN Women Bangladesh. The significance of this study on care systems lies in its potential to address the gap between the care supply and demand in Dhaka City. This study employs innovative methods to map care systems, including contextual factors that directly impact care demand, such as flooding, drought and epidemic risk. Dhaka City, the capital of Bangladesh, is of particular importance due to its role as the country's political and economic hub. It is the most densely populated area, facing unique challenges such as urban poverty and inadequate housing.

Bangladesh is structured into eight primary administrative divisions (Administrative Level 1), each further segmented into districts (Administrative Level 2), totalling 64 nationwide. The subsequent administrative tier consists of upazilas or thanas, amounting to 495 in total (Administrative Level 3). Following upazilas, there are union councils or wards (Administrative Level 4). Upazilas are pivotal in local governance, administration, and service delivery, serving as intermediaries between the central government and local communities. They are tasked with implementing government policies, coordinating local development efforts, and providing essential services to their respective populations.[1] This research examines Dhaka City at an 'area' level: a group of upazilas as its primary unit of analysis, as illustrated in Map 1.

Following the methodology proposed by De Los Santos (2022), this analysis focuses on identifying the demand and supply of care services in Dhaka. It aims to uncover potential discrepancies using an accessibility approach. Furthermore, the analysis of care needs (i.e., the demand side) integrates demographic and socioeconomic characteristics of the population (i.e., age, sex, poverty, labour force participation) as well as contextual factors, such as land use and environmental hazards. Multidimensional analysis not only highlights the value added of Geographic Information Systems (GIS) methods but also provides a deeper understanding of the care ecosystem in Bangladesh and better informs targeted interventions.

The next section outlines the steps for mapping care demand, incorporating contextual factors. Following this, the third section covers the process for mapping care supply. The fourth section details the strategy for assessing the accessibility of Dhaka's City care systems. Lastly, the fifth section addresses limitations and proposes mitigation strategies.

MAP 1. Area of Study



Source: Elaborated by the authors

II. Mapping Care Demand

Mapping the demand for care services involves identifying areas with a heightened need for such services based on the **demographic characteristics of the population**, with certain age groups usually associated with specific care provision. This report focuses on two populations: preschool-age age children (between 0 and 4 years old) and older adults (65 years old or older). Different potential variables can be included to analyse the demographic characteristics of these population groups, as shown in subsection A. Additionally, since literature has shown how contextual factors, like climate change-related events, land use or living conditions, impact care, these factors are analyzed and included, as presented in subsection B. These two subsections are divided into: i) data needs and data sources identification, ii) data preparation and cleaning, and iii) final indicators. Finally, subsection C presents how the areas for the analysis are constructed.

For both demographic and contextual factors, data from secondary sources is used. Each of the sources is presented in the corresponding subsection.

2.1 DEMOGRAPHIC VARIABLES

2.1.1 IDENTIFICATION OF THE NEEDS AND CORRESPONDING DATA SOURCES

In terms of demand for care services, data on population age distribution is needed. Once the types of data are defined it is necessary to address the data availability. Table 1 presents the sources used. Since the most recent census data publicly available is from 2011, which is now too outdated to accurately reflect the current population of Dhaka City; administrative data is not used. As an alternative, the estimates of the 2020 population by WorldPop are available. This data is used to present a more up-to-date population density.

TABLE 1. Care Demand Data Sources

Indicator	Type	Data source (with Link)	Date (last updated)	Granularity (lowest unit)	Details of the dataset
Population by Sex and Age Groups	Satellite Imagery	WorldPop	2020	~ 100 m	Estimates of the total number of people per grid square broken down by sex and age groupings (including 0-1 and by 5 years up to 80+). See the full list of available age groups here.

Source: Elaborated by the authors

2.1.2 DATA PREPARATION AND CLEANING

When using geospatial data, it is necessary to standardize the projection, resolution, and alignment of data. These tasks were managed using Quantum Geographical Information System (QGIS) version 3.36.3 with some performed in Python.

All rasters and shapefiles were processed in Dhaka's projection system, specifically UTM-46N (EPSG:32646).

The following preparation and cleaning are done for each kind of data:

- **[D2] Population by Sex and Age Groups:** WorldPop open data also includes the estimates of the total number of people per grid square broken down by sex and age groupings (including 0-1 years and by 5 years up to 80+) in 2020 for Bangladesh in units of number of women and men in each age group per grid square. These calculations are made based on the mapping approach by Pezzulo, C. et al. (2017)[2], sub-national mapping of population pyramids and dependency ratios in Africa and Asia.[3]

2.1.3 DEFINITION OF FINAL INDICATORS

Once the data is cleaned, the next crucial step is to define the indicators for analysis. In this context, identifying different age groups is essential because care needs vary significantly among them. Table 2 outlines the age group classification framework utilized in this research.

The final indicators used to assess the care demand are shown in Table 3.

TABLE 2. Age Group Division Framework for Care Demand Analysis

Age Group	Characteristics Type
Children aged 0 to 4 years old	Preschooling age
65+	Older adult population

Source: Elaborated by the authors

TABLE 3. Final Indicators for Care Demand Analysis

Indicator	Data source
Total population by upazila	WorldPop
0 - 4 year-old population by upazila	WorldPop
15 - 64 year old population by upazila	WorldPop
65 + year old population by upazila	WorldPop
Dependency ratio by upazila	WorldPop

Source: Elaborated by the authors

2.2 CONTEXTUAL FACTORS

Care is affected by contextual factors such as climate change and land use, which lead to different needs and priorities of specific populations. In Dhaka City, rapid urbanization, high population density, and the presence of over 5,000 slums exacerbate these challenges. Additionally, climate change adds a layer of complexity, as Bangladesh is vulnerable to extreme events, which can disrupt care systems and exacerbate vulnerabilities. Understanding these contextual factors is key to developing effective policies to support caregiving. This research is innovative by including contextual factors as key determinants of the care demand. For this, the following steps are done:

2.2.1 IDENTIFICATION OF THE DATA NEEDS AND DATA SOURCES

Three main types of factors were identified as necessary: disaster risk, landuse and proxies for socioeconomic distribution. Hereunder, the reasons are presented:

Disaster risk

The INFORM subnational model integrates various indicators related to **hazards, exposure, vulnerability** and coping capacity **at the upazila level** (Administrative Level 3).[4] It considers factors such as socioeconomic status, access to services, environmental risks, and demographic profiles, allowing planners to map the **distribution of risks and vulnerabilities across different Dhaka City**.

The research has a focus on floods due to several significant reasons:

1. **Dhaka City's susceptibility to flooding disproportionately affects vulnerable groups,** including women who often bear significant caregiving responsibilities.[5] Integrating flood risk assessments into care mapping helps prioritize resources for emergency response and evacuation, which is crucial for women managing caregiving duties amidst disrupted infrastructure.[6]
2. Floods can disrupt care services, affecting recipients and caregivers who rely on these facilities for their dependents. Integrating flood risk assessments into care mapping identifies critical care facilities needing fortification against floods.
3. **Flooding increases waterborne disease risks, which is particularly challenging for women managing sanitation and hygiene during crises.** [7]These health risks can have a detrimental effect not only on women but also on those they care for, as women often bear the primary caregiving responsibilities.
4. Understanding flood risk guides decisions on locations and designs for care facilities for children and older adults, aiming to minimize flood exposure and ensure resilient infrastructure.
5. Flood-focused care mapping contributes to **building community resilience**, as identifying vulnerable groups and areas can support proactive disaster preparedness, particularly among women who are central to caregiving roles.

Land Use

Urban and Peri-Urban Land Use/Land Cover (LULC) provides crucial information for analyzing and interpreting areas of varying density, **enhancing the contextualization of both the demand for and supply of care systems.** By identifying different land uses (e.g. residential, commercial),[8] **LULC helps identify areas with specific social needs, such as access to preschools, green spaces and recreational areas.** For care system strategies, understanding the distribution of different land uses can inform decisions on where to locate geriatric healthcare centres, preschools and daycare facilities, among others. For example:

- **Commercial and industrial units** refer to factories, warehouses, kilns, shopping malls and all associated facilities and land, including medium to large-scale compound units,[9] and are relevant for care supply in compliance with Bangladesh Labour Act 2006 requirements for childcare facilities in workplaces with 40 or more employees.[10] By providing onsite childcare, these units facilitate the integration of care responsibilities into daily work life.
- **Non-residential sites**, including leisure facilities like parks, serve as third spaces where caregivers can mobilize for childcare and eldercare activities. These areas provide opportunities for physical and social activities that are vital for the well-being of children and older adults. In particular, non-residential areas refer to governmental, medical, religious, military and educational areas, including associated areas.[11]

For a detailed definition of all the LULC, please see the table below.

Age Group	Characteristics Type
Formal high-density continuous residential (Sealing level > 80%)	Urban fabric is where formal housing (in the form of individual houses or apartment blocks) dominates. Formal houses are expected to be organized in a relatively regular spatial pattern with clearly visible roads. 0 %-20% of the urban fabric consists of non-sealed or vegetated surfaces.
Formal high-density discontinuous residential (Sealing level: 50% - 80%)	Urban fabric is where formal housing (in the form of individual houses or apartment blocks) dominates. Formal houses are expected to be organized in a relatively regular spatial pattern with clearly visible roads. 20 %-50% of the urban fabric consists of non-sealed or vegetated surfaces.
Formal low-density discontinuous residential (Sealing level: 10% - 50%)	Urban fabric consists of individual houses with little evidence of apartment blocks. Plot sizes are relatively large compared to formal high-density residential. 50% - 90% of the urban fabric consists of non-sealed or vegetated surfaces.
Village settlement (Sealing level < 10%)	Subclass of low-density residential class. Village settlements were labelled if located outside of the urban perimeter.
Commercial and industrial units	Factories, warehouses, kilns, shopping malls and all associated facilities and land, including medium to large scale compound units.
Non-residential urban fabric	Governmental, medical, religious, military and educational areas, including associated areas.
Roads and associated land	Motorways, primary and secondary roads.
Airports	Administrative area of airports, mostly fenced. Included are all airport installations: runways, buildings and associated land.
Construction sites	Areas with ongoing building/infrastructure construction activity or areas obviously prepared for construction.
Sand-filled area	Subclass of a construction site
Land without current use	Vacant land for which there is no evidence of ongoing building/infrastructure construction activity.
Urban greenery, sport and leisure facilities[12]	Urban green areas, recreational use as gardens, zoos and parks, including sports and leisure facilities and sports fields.
Agriculture	Cultivated areas, non-irrigated or permanently irrigated, including rice fields. Arable land (annual crops), permanent crops, orchards.
Other Natural and Semi-natural areas, including wetlands	Shrubs and/or herbaceous vegetation, including transitional woodland and wetlands.
Bare land	Natural areas where there is no or very little evidence of vegetation and do not serve as construction sites.
Bare land	Visible water surface areas.

Source: World Bank. (2023). Dhaka (Bangladesh) - Land Use / Land Cover Maps (ESA EO4SD-Urban).

The research has a special focus on **green spaces** because they are crucial for mitigating urban heat islands[13] and preserving precipitation patterns.[14] The presence of greenery and open spaces can significantly improve air quality, reduce stress, and offer a cooling effect, which is especially beneficial for young children and older adults, who are more susceptible to extreme temperatures.[15] Green Areas were estimated in the ESA EO4SD-Urban study using VHR Imagery (pan-sharpened multispectral), LULC product, OpenStreetMap (OSM) and include parks, City woodland, Sub-urban forest, Green alley and Sport & leisure facilities.[16]

Proxies for socioeconomic distribution

This research has a special focus on **informal settlements or slums[17]** due to several reasons:

- **Informal settlements often harbour a significant portion of the city's low-income population.** Examining these areas can provide insights into the economic disparities, living conditions and localization of the most vulnerable groups.
- **The inadequate quality and accessibility of basic services** in informal settlements highlight the socioeconomic challenges faced by residents. In Dhaka, these areas can reveal gaps in access to care systems.

- **Informal settlements are often associated with informal employment sectors.** Understanding the livelihoods in these areas can shed light on the economic landscape of the city, including the prevalence of informal labour.
  - In the case of unpaid care, it is highly likely that women who live in informal settlements are part of the informal economy.
  - Women in this scenario can face a triple workload (paid, unpaid and community), as well as a triple burden of constraints (being women, informal workers, members of poor households and disadvantaged communities).[18]
- **The spatial distribution of informal settlements can reveal patterns of urban growth and expansion,** which is especially important In Dhaka, where land use is highly contested.

Table 4 shows the data sources identified for contextual factors.

TABLE 4. Data Sources for Contextual Factors

Indicator	Type	Data source	Date[1]	Granularity[2]	Details of the dataset
Disaster Risk					
Multidimensional Risk Index	Geo Exposure -located index	European Commission[19]	2022	Upazila	Risk assessment is based on three key dimensions: hazard and exposure, vulnerability, and coping capacity limitations. Each dimension is divided into categories, such as natural disasters or infrastructure readiness, which are further detailed into components made up of specific indicators like earthquake risk, inequality levels, or healthcare availability.
Flood Extend and Hazard	VHR satellite imagery	ESA EO4SD-Urban[20]	2017	1x1 m	Flood extension and hazard for years 2004, 2007, 2014 and 2016
Land Use					
Land Use / Land Cover (LULC) type	VHR satellite imagery	ESA EO4SD-Urban	2017	1x1 m	In the urban context, the remote sensing-based LULC classes can be categorized into five Level 1 classes: Artificial areas, Natural/Semi-Natural, Agricultural, Wetland, and Water bodies.
Open & Green Areas coverage extent	VHR satellite imagery	ESA EO4SD-Urban	2017	1x1 m	In addition to the standard Urban Green Areas product offered by EO4SD, which includes two classes, an advanced service with a more detailed typology is provided: Open and Green Spaces (OGS).
Proxies for socioeconomic distribution					
Informal Settlements	VHR satellite imagery	Informal Settlements (ESA EO4SD-Urba)	2010/ 2017	0.1 – 0.25 ha (MMU)	Probable informal settlements over Dhaka contain spatial explicit information about the position of slums as identified in 2006 and 2010 from ancillary data sources and in 2017 by interpretation of VHR satellite imagery.

1 Last updated
2 Lowest unit

Source: Elaborated by the authors

2.2.2 DATA PREPARATION AND CLEANING

Similar to 2.A.2, data with pixel values are resampled to upazilas using zonal statistics. This technique involves aggregating pixel-level data within specified geographic boundaries to obtain representative values for each upazila. For instance, in the case of informal settlements, land use, and open green areas, each pixel within the dataset is assigned a binary value: 1 if a specific characteristic is present and 0 if it is absent. To calculate the total area covered by a particular characteristic within an upazila, the values of all relevant pixels are summed. This sum is then multiplied by the pixel area (1m x 1m), providing an estimate of the total area occupied by that characteristic.

The flood extent information is derived from historical optical satellite imagery with a resolution of 1mx1m. This data offers a spatial representation of areas affected by flooding over time. On the other hand, flood hazard information is generated by analyzing the occurrence of flood events over the past 10 years, taking into account both recent and significant historical events. Notably, the analysis includes the catastrophic flood event of 2004, which serves as a key reference point. By combining these historical records with recent flooding patterns, a comprehensive assessment of flood hazards is developed, enabling more accurate risk analysis and mitigation planning.

Finally, the INFORM risk index comes at an upazila level from the original source.

2.2.3 FINAL INDICATORS

TABLE 5. Final Indicators for Contextual Factors

Indicator	Source
Disaster Risk	
Area (in km2) of flood extent per Upazila in 2004	ESA EO4SD-Urban
Area (in km2) of flood extent per Upazila in 2007	
Area (in km2) of flood extent per Upazila in 2014	
Area (in km2) of flood extent per Upazila in 2016	
Total area (in km2) of flood extent per Upazila in years 2004, 2007, 2014, 2016	
Area (in km2) hit by 1 flood hazard event per Upazila in years 2004, 2007, 2014, 2016	
Area (in km2) hit by 2 flood hazard events per Upazila in years 2004, 2007, 2014, 2016	
Area (in km2) hit by 3 flood hazard events per Upazila in years 2004, 2007, 2014, 2016	
Area (in km2) hit by 4 flood hazard events per Upazila in years 2004, 2007, 2014, 2016	
Share (%) of high flood risk area among total surface per Upazila	
Multidimensional Risk Index per Upazila	European Commission
Drought Risk per Upazila	
Flood Risk per Upazila	
Epidemic Risk per Upazila	

Indicator		Source
Land Use		
Area (in km2) of Mine, Dump and Construction Sites per Upazila in 2017	Share (%) of Mine, Dump and Construction Sites per Upazila in 2017	ESA EO4SD-Urban
Area (in km2) of Industrial, Commercial, Public, Military, Private and Transport Units per Upazila in 2017	Share (%) of Industrial, Commercial, Public, Military, Private and Transport Units per Upazila in 2017	
Area (in km2) of Urban Fabric per Upazila in 2017	Share (%) of Urban Fabric per Upazila in 2017	
Area (in km2) of Bare land per Upazila in 2017	Share (%) of Bare land per Upazila in 2017	
Area (in km2) of Agricultural Area per Upazila in 2017	Share (%) of Agricultural Area per Upazila in 2017	
Area (in km2) of Artificial non-agricultural vegetated areas per Upazila in 2017	Share (%) of Artificial non-agricultural vegetated areas per Upazila in 2017	
Area (in km2) of Forest per Upazila in 2017	Share (%) of Forest per Upazila in 2017	
Area (in km2) of Other Natural and Semi-natural Areas (Savannah, Grassland) per Upazila in 2017	Share (%) of Other Natural and Semi-natural Areas (Savannah, Grassland) per Upazila in 2017	
Area (in km2) of Inland Water per Upazila in 2017	Share (%) of Inland Water per Upazila in 2017	
Area (in km2) of open and green areas per Upazila in 2017	Share (%) of open and green areas among total surface area per Upazila in 2017	
Proxies for socioeconomic distribution		
Area (in km2) of informal settlements per Upazila in 2017	Share (%) informal settlements among total surface per Upazila in 2017	ESA EO4SD-Urban

Source: Elaborated by the authors

2.3 IDENTIFYING MAIN AREAS: CLUSTERING MAIN VARIABLES

To simplify the analysis, different demographic and contextual factors variables are used to group the upazilas in **areas**. The following criteria were used for this purpose:

- **Geographic Contiguity:** Ensuring that the upazilas within each cluster are geographically contiguous for efficient administration.
- **Dhaka City Corporation:** Belonging to the same administrative corporation, Dhaka North City Corporation (DNCC) or Dhaka South City Corporation (DSCC).[21] These are the administrative bodies overseeing Dhaka's wards, which are tasked with delivering essential urban services.[22]
- **Functional Roles:** Grouping upazilas with similar functional roles to streamline urban planning efforts.
- **Geographic Characteristics:** Ensuring each cluster considers the unique needs of each area, including flooding hazards, green areas, residential areas, etc.
- **Socioeconomic level:** Considering socioeconomic integration within clusters, including overall living standards and economic opportunities, and informal settlements shared by upazila.

A hierarchical clustering was implemented. The resulting clusters were manually verified based on qualitative literature. The indicators from Table 6 were used for the model.

TABLE 6. Indicators for Areas

Indicator	Source
Demographic characteristics	
Dependency ratio by upazila	Self-calculation based on WorldPop
0 - 4 year-old population by upazila	WorldPop
65 + year old population by upazila	
Total population by upazila	
Disaster Risk	
Total area (in km2) of flood extent per Upazila in years 2004, 2007, 2014, 2016	ESA EO4SD-Urban
Land Use	
Share (%) of Mine, Dump and Construction Sites per Upazila in 2017	ESA EO4SD-Urban
Share (%) of Industrial, Commercial, Public, Military, Private and Transport Units per Upazila in 2017	
Share (%) of Urban Fabric per Upazila in 2017	
Share (%) of Bare land per Upazila in 2017	
Share (%) of Agricultural Area per Upazila in 2017	
Share (%) of Artificial non-agricultural vegetated areas per Upazila in 2017	
Share (%) of Forest per Upazila in 2017	
Share (%) of Other Natural and Semi-natural Areas (Savannah, Grassland) per Upazila in 2017	
Share (%) of Inland Water per Upazila in 2017	
Share (%) of open and green areas among total surface area per Upazila in 2017	
Proxies for socioeconomic distribution	
Mean Relative Wealth Index per Upazila	Data4good (Meta)
Share (%) informal settlements among total surface per Upazila in 2017	ESA EO4SD-Urban

Source: Elaborated by the authors

Considering these criteria, Map 1 presents the resulting seven areas, four of them located in DNCC and three of them in DSCC.

III. Mapping Care Supply

Mapping the care supply implies identifying **areas with service provision for children of preschool age and older adults.** This subsection presents the four steps needed for mapping care supply: i) Identification of the data needs and data sources, ii) data preparation and iii) definition of final indicators.

The report uses three main types of data for care supply:

- **Administrative geo-referenced data** refers to statistical information collected and maintained by governmental or official agencies, typically derived from sources like population censuses, household surveys, or administrative records. These datasets are geo-referenced, meaning they include location information (latitude and longitude) tied to specific administrative boundaries such as districts, municipalities, or other administrative divisions. They provide structured, authoritative information used for planning, policy-making, and geographic analysis, although they may not always be updated in real-time or have detailed information at smaller administrative levels.
- **Crowdsourced data,** particularly from platforms like OpenStreetMap (OSM), is collected collaboratively by volunteers and users worldwide. It involves the community-driven mapping of geographic features, including roads, buildings, amenities, and points of interest. Unlike administrative data, which is typically managed by governmental bodies, crowdsourced data on OSM is open and continuously updated by contributors globally. This data is valuable for its real-time updates, diverse coverage (including informal settlements and local amenities), and flexibility in adding new features or correcting existing ones. However, it may vary in accuracy and completeness depending on local community engagement and verification processes.
- **Google Places API** - Web scraping and data preprocessing involve extracting information from websites and providing access to care facility lists not found in other data sources. For older adult care facilities, initial desk research found no secondary geo-located sources, necessitating web scraping of Google Maps.

3.1 IDENTIFICATION OF THE DATA NEEDS AND DATA SOURCES

In terms of supply, four types of data are identified:

- Childcare Facilities
- Older Adult Care Facilities
- Education Facilities
- Health Facilities

This research focuses primarily on preschool childcare and older adult care. However, it's essential to acknowledge that education facilities, particularly those serving preschool-aged children (such as kindergartens), play a critical role in childcare by providing a safe and nurturing environment during working hours. Similarly, healthcare facilities are integral to older adult care, as the literature consistently highlights that a significant portion of older adult care services is centred on health-related needs.

To gain a comprehensive understanding of the landscape of childcare and older adult care, it is crucial to include data on both education and health facilities. These datasets are compiled and cleaned separately from the primary categories of childcare and older adult care facilities. After thorough processing, the education and health data are integrated with the respective childcare and older adult care data sets. This holistic approach ensures that all relevant aspects of care are considered, providing a complete picture of the available resources and the services they offer. Table 7 presents the sources identified for mapping the care supply. Google API is based on a list of keywords presented in Table 8.

TABLE 7. Care Supply Data Sources

Indicator	Type	Data source	Date[1]	Granularity[2]	Details
Childcare Facilities	Google API	Google API	2024	Point data	Google Places API is a product the Google Maps Platform offers through Google Cloud Console. The API allows users to interact with the platform using specialized libraries, in this case, the library Googlemaps from Python 3.12.3. The places API includes information about establishments, geographic locations, and general points of interest such as status, name, geometry, ratings and number of raters. For this research, a set of keywords was defined by Data-Pop Alliance and UN Women (see Table 8)
Older Adult Care Facilities	Google API	Google API	2024	Point data	
Education Facilities	Crowdsourced Geo-located data	Humanitarian OpenStreetMap (HOT)	2024	Point data	In the urban context, the remote sensing-based LULC classes can be categorized into five Level 1 classes: Artificial areas, Natural/Semi-Natural, Agricultural, Wetland, and Water bodies.
	Geo-located Administrative data	Provided by UN Women	2024	Polygon data	In addition to the standard Urban Green Areas product offered by EO4SD, which includes two classes, an advanced service with a more detailed typology is provided: Open and Green Spaces (OGS).
Health Facilities	Crowdsourced Geo-located data	Global Health Sites Mapping Project	2024	Point data	The data from the Global Health Sites Mapping Project provides information about four types of amenities or buildings. The dataset includes information on the name, address, and unique ID, among other variables.
	Geo-located Administrative data	Provided by UN Women	2024	Polygon data	See Education - Geo-located Administrative data.

1 Last updated
2 Lowest unit

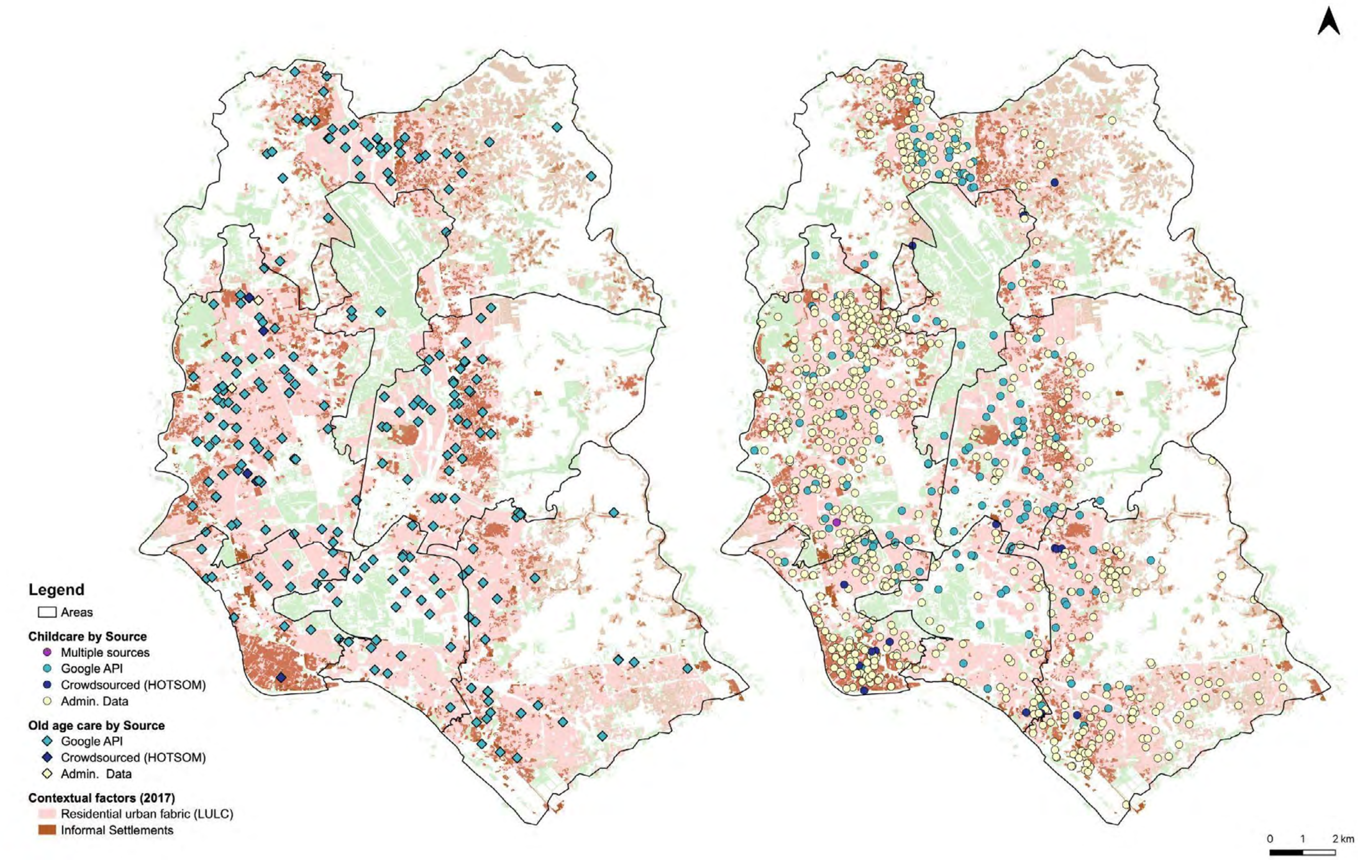
Source: Elaborated by the authors

TABLE 8. List of Keywords

Indicator	Type	Data source	Date1
		English	Bengali
Childcare	Childcare in preschooling age (0-5 years old)	Daycares	দিবায Cক
		Child Daycare	শি দিবায Cক
		Private daycares	Cবসরকারি শি দিবায Cক
		Nurseries	নাস ারি, শি শালা
		Nursery schools	নাস ারি ুল
		Crèches	নাস ারি
		Preschools	াক বিদ7ালয়
		Kindergartens	কিVারগাটে ন
		Community-based childcare centres	সমাজভিি ক শি -য Cক
		Childcare cooperatives	শি য সমবায়
		Employer-sponsored childcare	নিয়োগকারী তি ানের শি -য Cক
		Garment factory childcare facilities	Zতরি Cপাশাক কারখানার শি য ব7বা
		Child development centres	শি বিকাশ Cক
		Early childhood education centres	শি র ারি ক শিWা Cক, াক-াথমিক শিWা Cক
		Nanny	আয়া, শি পালনকারী
		Early learning centres	ারি ক শিWা Cক, াথমিক শিWা Cক
Older adult care 65+	Non-health related	Assisted living facilities	আসবাবপ স লিত বসবাসের সুবিধা
		Nursing homes	চিকিৎসালয়, হাসপাতাল
		Long-term care	দীঘ ায়ী Cসবা
		Retirement communities	অবসর া9 স দায়, অবসর া9 ব7বিগ
		Continuing care retirement communities (CCRCs)	অবসর া9 স দায়ের সাব Wনিক Cসবা
		Senior living apartments	বয় ব7াদের বাসান
		Adult daycare centres	বয় ব7াদের দিবায Cক
		Older adult caregivers	বয় বাদের পরিচয াকারী
		Old home	বd াম
		Older adult care cooperatives	বয় য সমবায়
	Health related	Geriatric hospitals	বাধ ক7 হাসপাতাল
		Midwife	ধা ী
		Nurse	Cসবিকা, নাস
		Community Health Worker	কমিউনিP Wা7কম , ানীয় Wাকম
		Geriatric facilities	বয় Wা7সেবা সুবিধা
		Skilled nursing facilities	দW নাসি ং সুবিধা
		Alzheimer's care homes	আলজাইমার Cসবা Cক
		Hospice care	ধমশালা য কে
		Senior rehabilitation centres	বয় পুনব াসন Cক

Source: Elaborated by the authors

MAP 2. Distribution of Old-Age Care Facilities (left) and Childcare Facilities (right) by data source



Source: Elaborated by the authors

3.2 DATA PREPARATION AND CLEANING

The data preparation and cleaning process follows similar steps for each data source. However, due to the varying structures of these sources, the specific steps for data preparation and cleaning are detailed individually for each data type.

3.2.1 GOOGLE PLACES API: CHILDCARE FACILITIES, OLDER ADULT CARE FACILITIES, EDUCATION FACILITIES, HEALTH FACILITIES

Step 1: Extraction

The API supports multiple types of place searches, including “nearby search” and “text search”. The nearby search lets users find places within a certain radius, whereas text search provides results based on a search string. In this case, the nearby search was used to find the places of interest, defined with a set of keywords, around certain coordinates.

Through the Nearby Search function of Google Places API it is possible to locate places by use of a given set of search terms. In this case, the API uses some user-provided keywords and identifies places within a certain radius from a given point. The search results are based on how well the details provided about the establishments, their names, and related metadata match with the entered keywords.

Using the API has certain limitations in its search approach. The metadata contains details such as the name, location, category and user reviews of a place. The API uses these data points for keyword searches, which means that slight differences in how places are described or classified may lead to irrelevant search results. The accuracy of the results depends heavily on how well the keywords match the places' metadata, which may not always be comprehensive or correctly tagged, leading to incomplete or irrelevant results. Moreover, a keyword-based approach might exclude places that do not directly mention keyword phrases in their metadata, but they are still relevant for users' intentions behind querying thus leading to biased search results.

One of the main drawbacks of the Google Places API is that each query gives a maximum of sixty places. Therefore, if the radius chosen in the nearby search is too big, an important number of places may not appear in the search. To overcome this limitation, a rectangular grid was produced covering Dhaka City, with a distance of 1.5 km between points, obtaining a total of 280 points. Some of them do not belong to the city and are filtered later, but they are produced in the grid for the convenience of making it rectangular. Next, a query is made at each point to obtain places of interest within a radius of 1 km. With this combination of values of 1.5 km between points and a 1 km radius, it is ensured that the circles centred at each point with a radius of 1 km cover the entire city, with several overlaps between them but no gaps. This method is chosen because it is better to have duplicates in the dataset, which can be filtered during cleaning, rather than losing places of interest due to gaps in the grid.

Once the data for each point is collected, locations with exactly 60 places of interest are reanalyzed, as they may actually contain more but were capped at 60 due to platform limitations. To address this, a smaller local grid is created, and additional queries are performed at these new grid points, ensuring that all places of interest in the area are captured.

The dataset resulting from this method contains many duplicates, as places belonging to the overlapping regions in the grid will appear at least twice. This makes cleaning the dataset of vital importance.

Step 2: Deduplication

Three criteria are used for deduplication. First, the Google API facilities come with a unique key called 'place_id', and only one observation per place_id is retained. Second, latitude and longitude are used to identify potential duplicates within defined radio. The first radius involves rounding to six decimals, equivalent to 11.1 cm or approximately 4 inches. The second radius involves rounding to five decimals, equivalent to 1.11 meters, allowing for differentiation between individual trees. A script was written to identify facility names within these distances, and human judgment was used to determine if the two facilities were the same based on their names. If the names are similar, a Google search is conducted to verify the existence of two distinct facilities. If the names are different, both facilities are retained. Third, facilities come with a 'type' variable, and some types are considered suspicious for the project's scope. For example, childcare facilities may be listed under types like 'clothing store'. These suspicious types are placed into sub-datasets and manually verified. Verification is based on the facility's name; if the name does not clearly indicate a relationship to care, a Google search is performed to verify the facility's purpose.

3.2.2 CROWDSOURCED DATA: EDUCATION FACILITIES AND HEALTH FACILITIES

Step 1: Extraction and categorization

The crowdsourced data from Humanitarian OpenStreetMap includes an "amenity" variable with five categories for all Bangladesh:

- College - 530
- Kindergarten - 147
- Marketplace - 1
- School - 5153
- University - 143
- NA - 109

Since this research is focused on preschool childcare, only kindergarten and NAs are kept. Data is filtered to keep the facilities only within the study area.

Regarding healthcare, data from the Global Health Sites Mapping project includes an "amenity" variable:

- Bridge - 1
- Clinic - 759
- Dentist - 199
- Doctors - 116
- Hospital - 1270
- Laboratory - 1
- Pharmacy - 3053
- NA - 108

UN Women's prioritization guided the filtering to retain "clinic," "doctors," and "hospital" types. NAs are kept for further analysis. Data is filtered to keep the observations within the area of study.

Step 2: Data Completion

The NAs are completed based on the information from the 'building' variable when possible. 'Building' can include information about the type of usage that takes place in the building. When 'kindergarten', 'hospital' or 'doctor' is found in 'building', the NA under 'amenity' is replaced. Since the datasource does not come with latitude and longitude but with an 'object_id' variable, latitude and longitude are added using QGis.

Step 3: Deduplication

The observations are grouped by similar names using the 'fuzzy' function from Python and a 111-meter radius. A manual inspection is done based on the name and address to define if the two facilities are or are not the same.

Step 4: Filtering

The observations are filtered on QGis to keep only the ones within the study area.

Step 5: Older adult care identification

A manual search is done to identify healthcare facilities focused on older adult care. Facilities that include in their name: geriatric, midwife, nurse, community health worker, nursing, Alzheimer's, hospice or senior are kept.

3.2.3 ADMINISTRATIVE DATA: EDUCATION FACILITIES AND HEALTH FACILITIES

A. Education Facilities

Step 1: Extraction and categorization

Administrative data includes two main categories: 'Education and Research' and 'Health Facilities'. The first step involves extracting and categorizing the data using a predefined list of keywords. For the 'Education and Research' category, administrative data includes information on six groups:

- Kindergarten and Nursery
- Pre-primary School
- Primary School
- Madrasa (Pre-primary)
- Madrasa (Primary to High School)
- High School

Since this research is focused on preschooling childcare, data is filtered to keep kindergarten and nursery, pre-primary school and madrasa (pre-primary level).

Step 2: Data Completion

The variable "Str_3" includes "NA". The NAs are replaced with the information from floor usage when possible.

Step 3: Deduplication

The primary objective of the deduplication process is to keep a single point or polygon for each place. This necessity arises due to multiple records that may exist for the same building, each with varying levels of information. Key information required for this deduplication process includes the building name and location, such as the address or spatial coordinates. The process involves three steps to address exact matches, typographical errors, and missing values.

In the first step, buildings with exact matches in both name and address are identified. For buildings with complete (non-null) names and addresses, duplicates are identified and removed if both the name and address are identical. The cleaned dataset from this step is then combined with records that were not processed.

The second step focuses on buildings with similar names that may have slight variations but are within a close spatial distance using the Fuzzy function. Centroids from the polygon layer are extracted to aid in this analysis. For buildings with non-null names, we retain only one building from groups of similar names within 0.001 degrees (~111 meters). After a manual review to ensure accuracy, the cleaned dataset is combined with the unprocessed records from this step.

The final step deals with buildings that have missing values in the name or address fields. Centroids that are within 0.0001 degrees (~11 meters) and belong to the same category are compared. In these cases, the building with less information is removed, ensuring that the most complete dataset is retained.

Step 4: Merge

The clean datasets of Dhaka North and Dhaka South are merged.

B. Older adult Facilities

Step 1: Extraction and Categorization

To identify records specifically related to older adult care and exclude general medical services such as doctors, clinics, and hospitals, a targeted keyword matching approach was employed. The predefined keyword list was used to search the administrative data, isolating records containing the specified keywords. Facilities with: geriatric, midwives, nurses, community health workers, nursing, Alzheimer's, hospice or seniors are kept.

Step 2: Merge

The identified facilities from the Dhaka North and Dhaka South datasets are merged.

3.3 DATA INTEGRATION

3.3.1 CHILDCARE INTEGRATION

To keep only one master dataset for childcare, data sources from childcare facilities and education facilities focused on preschool age are merged.

Step 1: Data preprocessing

The necessary columns are extracted from the different data sources. Specifically, name, address, source, latitude, and longitude. The column names are standardized to ensure uniformity across all datasets. This step is essential for maintaining consistency and facilitating the merging of data from different sources.

Step 2: Data integration for Google API and Administrative Data

For integrating the Google API and Administrative datasets, a spatial proximity criterion to find potential matches is used, considering two locations as potential matches if they are within 0.005 degrees (~555 meters). The 0.005-degree threshold (~555 meters) was chosen to balance the need for spatial accuracy with the inherent imprecision of geographic coordinates, ensuring relevant proximity without being overly restrictive.

Once the potential matches are identified the name is used to define if the two observations are or not the same. No matches are found. It must be noted that this approach is limited because multiple observations from administrative data have an empty name. When this happens, the two facilities are kept to avoid underestimating the number of facilities.

Two columns are created after identifying the duplicates: is_google and is_admin. These dummy variables take a value of 0 if a facility is not in the respective data source and 1 if it is. This data structure ensures that each facility has a single row, allowing for the identification of duplicate data sources.

Step 3: Integration of HOTSOM Data

Next, the HOTSOM dataset with the preliminary dataset is integrated using a similar approach. Each HOTSOM location is compared to the ones in the merged data set using the 0.005 degrees criteria. Then, the names are used to identify if the two observations are or are not the same. Since multiple observations do not have a name, this is a limited approach. When a named entity is matched with an unnamed entity, the two are left as different entities. Therefore, this approach might be overcounting the facilities. The dummy variable 'is_hotsom' is added.

Step 4: Identification of public facilities

An additional column is added to identify facilities that are public based on their name. A search of the terms 'public', 'official', and 'government' is done. If a facility includes any of these words in the name, it is considered public. Thirteen matches are found.

Step 5: Type division

In order to divide the preschool-age facilities identified and given the data limitation, the following decisions are made:

- Administrative data is the one with the highest number of facilities. This includes kindergarten and nursery, pre-primary school and pre-primary level madrasa. These categories are kept.
- Crowdsourced data only includes 'kindergarten'. All observations from this datasource are added to 'kindergarten&nurseries'.
- Google API does not include any variable to identify the kind of facility. Keywords 'kindergarten' and 'nursery' are looked for in the names, and the matching results are added to 'kindergarten&nurseries'. 'Madrasa' is looked for, and the matching results are added to 'pre-primary level madrasa'. Matching results for 'primary' and 'preschool' are added to 'pre-primary school'. Finally, 'school' is added to 'kindergarten&nursery'. This decision is taken after verifying that administrative data includes multiple 'schools' in this category. Finally, all the non-classified facilities are manually verified to be reclassified. If they can not be integrated into a different category, they are considered 'other'.

3.3.2 OLDER ADULT CARE AND HEALTHCARE INTEGRATION [S2] AND [S4]

The primary data source for older adult care is Google. Based on specific keywords, the Google-listed facilities are categorized as either 'is_health' or 'is_non_health' (see Table 8). In the crowdsourced data, five facilities include one of the keywords related to older adult care. The names of these facilities are manually searched within the Google dataset. Since none of these facilities are found in Google, all of them are added to the dataset. A manual review is then conducted to determine whether the nursing facilities are health-related, and they are classified accordingly.

A similar procedure is followed with administrative data. Two facilities in this dataset have names that include keywords specific to older adult healthcare facilities. These names are compared against the existing dataset, and the facilities are added as needed.

3.3.3 SUMMARY OF THE INTEGRATION STRATEGY

Table 9 presents a summary of the content of the original data sources and the filtering strategies to merge the different datasets. The main goal is to keep relevant information only for pre-primary childcare and older adult care.

Table 9. Filtering Strategy Summary

Source	Childcare Facilities	Older Adult Care Facilities	Education Facilities	Health Facilities
Google API	**Original**			
	Keywords for preschooling age (See Table 8) Filtering strategy	19 keywords grouped in Non-health-related facilities, Health-related facilities Filtering strategy	Keywords for preschooling age (See Table 8)	Three keywords: Clinic, Hospital, Doctors,
	Filtering strategy			
	Keywords for preschooling age (See Table 8)	Keywords related to older adult health-focused facilities vs. non-focused (See Table 8)	Keywords for preschooling age (See Table 8)	Keywords related to older adult-health- focused facilities: Geriatric, midwife, nurse, community health worker, nursing, Alzheimer's, hospice or senior.
Crowdsourced data	**Original**			
	NA	NA	Amenity: Kindergarten, School	Amenity: Clinic, Hospital, Doctors,
	Filtering strategy			
	NA	NA	Kindergarten	Keywords related to older adult-health -focused facilities: Geriatric, midwife, nurse, community health worker, nursing, Alzheimer's, hospice or senior.
Administrative data	**Original**			
	NA	NA	Edu_o	Str_Use2: Community Health, Hospital
	Filtering strategy			
	NA	NA	Kindergarten	Keywords related to older adult-health-focused facilities: Geriatric, midwife, nurse, community health worker, nursing, Alzheimer's, hospice or senior.

Source: Elaborated by the authors

3.4 FINAL INDICATORS

Table 10 presents the content of the final childcare and older adult care datasources. For childcare services, the data allows for categorization into five distinct groups. Data on older adult care facilities is considerably less detailed, allowing only a broad categorization into two groups. This lack of granularity in older adult care data points to a broader issue of insufficient data collection and reporting focused on older adults, which hinders the ability to fully understand the specific needs of this population.

TABLE 10. Final Datasets

	Preschool care1					Older adult care	
	Kindergarten2, Nursery	Pre-primary school	Madrasa	Daycare	Other	Health-related	Non-health- related
Number of facilities	735	6	19	78	44	68	165
Total	873					233	

1 The reason why adding the different types results in 9 “extra” facilities is because the categorization of the Google API data is done based on the facility´s name. The name can include two keywords, such as “XYZ School and Daycare”. In this case, the facility is included both in kindergarten, nursery and school, and in daycare.

2 Includes kindergartens that only offer this educational level and schools that offer kindergarten

Source: Elaborated by the authors

IV. Assessing Dhaka's City Care Systems

4.1 DATA INPUTS, PARAMETERS AND ASSUMPTIONS

AccessMod 5 is a software tool developed by the World Health Organization (WHO) to assess accessibility to health services, focusing on geographic accessibility and service availability. The tool has evolved with contributions from various organizations, including the WHO Department of Health Systems Governance and Financing, the University of Geneva's Institute of Global Health, the AeHIN GIS Lab, and the WHO eHealth unit.

To model accessibility to care facilities in Dhaka City, factors such as topography, road networks, movement constraints (e.g., rivers, lakes, flood extents), target population distribution, and the locations of care facilities are considered (see Table 11).

TABLE 11. Final Data Inputs for AccessMod

Indicator	Type	Data source	Date	Resolution	Uses
	Other inputs				
Childcare facilities	Vector	Merged datasets	Multiple	Facility location	
Older ault care facilities	Vector	Merged datasets	Multiple	Facility location	
	Other inputs				
Digital Elevation Model (DEM)	Raster	Copernicus DEM	2023	30m	Used to run anisotropic analysis (slopes would affect the speed of travel)
Road network	Vector	Humanitarian OpenStreetMap (HOT)	2024	Roads	Used to extend much further outward the facility as travel time could be faster.
Population distribution grid	Raster	WorldPop	2020	100m	

Indicator	Type	Data source	Date	Resolution	Uses
Land cover distribution grid /	Raster	Land Use / Land Cover Maps (ESA EO4SD-Urban)	2017	Core Urban Area: 0.5 m	Used to characterize the area of analysis and that are believed to affect the travelling time of patients moving across this area.
Barriers to movement	Raster	Land Use / Land Cover Maps (ESA EO4SD-Urban)	2017	Core Urban Area: 0.5 m	Used to represent various components of the landscape the population cannot travel through, such as rivers, lakes, military and airport zones and industrial complexes.
Exclusion areas	Raster	Land Use / Land Cover Maps (ESA EO4SD-Urban)	2017	Core Urban Area: 0.5 m	Used to consider areas where a share of the target population might be living but where facilities could/ should not be located. *Examples of such areas may include low-lying coastal* *areas subject to inundation/ storm, national parks, military zones, etc.*
Zones boundaries	Vector	OCHA	2024	Upazila/ Districts	

Source: Elaborated by the authors

To apply AccessMod 5 for accessibility analysis, it is necessary to consider several input parameters. These parameters are essential for capturing local needs and traffic conditions accurately and, therefore, ensuring that the analysis results are both precise and contextually relevant.

The assumptions guiding the accessibility modelling are as follows: (i) the maximum travel time of 15 minutes reflects a practical limit based on empirical evidence for accessing care facilities; (ii) travel is predominantly motorized on Roads, with average speeds set at 8 km/h for primary roads, 6 km/h for secondary roads, and 4 km/h for tertiary roads, reflecting local vehicular conditions in Dhaka; and (iii) walking is the primary mode of transport on land covers, with an assumed average walking speed of 3 km/h based on limited regional studies.

4.1.1 MAXIMUM TRAVEL TIME

Firstly, the **maximum travel time** is a critical parameter that defines the catchment area for each facility. This parameter represents the maximum duration a child or older adult person can feasibly travel to access a care facility, contingent upon the severity of their condition. The determination of the maximum travel time requires rigorous analysis and must be based on empirical evidence to ensure it accurately reflects realistic scenarios.

Current research on this topic primarily focuses on summarizing travel statistics rather than establishing a universally acceptable travel time limit. For example, the 2021 Travel Use Survey by the Bangladesh Bureau of Statistics and UN Women Bangladesh reports that the average time spent on activities such as travelling, moving, transporting, or accompanying goods or persons related to unpaid domestic services is 5.4 minutes. Additionally, a 2018 study on healthcare accessibility for older adult individuals in Bangladesh found that 68.8% of respondents reported travel times between 15 and 30 minutes. Based on these findings, we have determined that an acceptable maximum travel time is 15 minutes.

4.1.2 Travel Scenario Table

Additionally, the mode and speed of travel play a significant role in the results of AccessMod analyses. The travel scenario table in AccessMod allows for the assignment of different travel modes (walking, motorized, and bicycling) and corresponding speeds to each category. The selection of these travel parameters must be grounded in robust empirical data and relevant studies to accurately represent the conditions encountered by the target population.

Initially, we assume that travel on all road categories is predominantly by vehicle, including vehicles such as automobiles, buses, non-auto rickshaws and bicycles. According to the World Bank's 2022 Integrated Corridor Management Dhaka North Project, the average vehicular speed in Dhaka is approximately 6 km/h. This finding is corroborated by a 2022 local media report indicating that the average vehicle speed in Dhaka has decreased to 4.8 km/h. Based on this data, we propose setting the average speeds for primary, secondary, and tertiary roads at 8 km/h, 6 km/h, and 4 km/h, respectively.

For various land use types, we assume that the primary mode of travel for the target population is walking. A 2012 study found that the average free-flow walking speeds for children and older adult individuals in Dhaka City are 3.7 km/h and 3.8 km/h, respectively. Although there is limited literature measuring actual walking speeds in Dhaka City, research from Khulna City, the third-largest city in Bangladesh, suggests a mean walking speed of 3.1 km/h. Given the minimal variance in walking speeds between children, older adults, and the general population (less than 0.25 km/h), we assume a walking speed of 3 km/h for both children and older adult individuals.

4.1.3 BARRIERS

Certain restricted zones must be excluded from the analysis. For example, Dhaka includes 2,030 hectares of restricted areas occupied by military or government establishments, which are not accessible to the public. Key locations within these restricted zones include the President's Office and Residence at Banga Bhaban, the Prime Minister's Office and Residence, the National Parliament Building, Hazrat Shahjalal International Airport, the Secretariat near Paltan, and the Old Airport Area at Tejgaon.

4.2 DATA PROCESSING IN ACCESSMOD

Before conducting the main analysis in AccessMod, several preprocessing steps are required to ensure the accuracy and reliability of the results. The following sections introduce these tasks: filtering facilities on barrier pixels, adjusting population data, and merging land cover information.

4.2.1 POPULATION ADJUSTMENT

This step refines the population distribution by redistributing data from barrier pixels—areas blocked by features like roads or rivers—to nearby regions that are habitable. This adjustment provides a more accurate representation of population coverage and catchment areas.

4.2.2 LAND COVER MERGING

Different layers—such as the road network, water bodies, and existing land cover—are combined into a single, unified land cover layer. The resulting layer gives a detailed view of Dhaka City's landscape and supports further pixel-based analysis in AccesMod.

4.2.3 FACILITY FILTERING ON BARRIERS

In this step, facilities incorrectly placed on barrier pixels, such as roads or rivers, are identified and removed. This ensures that the analysis only includes facilities that are realistically accessible, preventing any distortion of results caused by impractical facility locations.

4.3 TYPES OF ANALYSES

The following sections detail various analytical approaches used to assess and enhance the accessibility and coverage of care facilities. Each type of analysis leverages spatial data and specialized tools to provide insights into the distribution and availability of care systems across different regions.

4.3.1 ACCESSIBILITY ASSESSMENT

The Accessibility Analysis in AccessMod is designed to compute the travel time surface, determining the time required to reach the nearest facility. This analysis incorporates land cover data, facility locations, maximum travel time, and a travel scenario table that specifies travel speeds for each combined land cover class. The resulting output is a raster-format travel time layer, providing a highly detailed representation of accessibility to facilities.

4.3.2 GEOGRAPHIC COVERAGE ANALYSIS

The geographic coverage analysis evaluates the ability of health facilities to provide care to the target population, considering both their geographic accessibility and capacity. In addition to inputs from the accessibility assessment, this analysis incorporates population data and the maximum capacity of each facility to estimate the portion of the population that remains underserved, even if they are physically within reach. The primary outputs include the spatial distribution of the residual population, presented in raster format, and the catchment areas for each facility, provided in vector format. This analysis can also be performed without considering facility capacities, in which case the results are only constrained by the maximum travel time.

4.3.3 ZONAL STATISTICS

Zonal Statistics determines the percentage of the population covered within each sub-national division. By combining the travel time distribution grid—generated using the "accessibility analysis" tool—with the spatial distribution of the target population and the boundaries of the zones, the analysis provides a detailed assessment of coverage. The output is a table that includes the total population, the covered population, and the percentage of the population that is covered within each zone.

V. Limitations

5.1 GENERAL LIMITATIONS

Bias towards education: A major limitation of this research is the absence of a comprehensive, structured dataset encompassing all existing childcare facilities. This results in a bias towards kindergartens since facilities not related to education, such as daycares, are only identifiable through Google API searches. Consequently, many childcare facilities, including daycares and after-school programs, may be underrepresented in the analysis.

Lack of older adult data: The identification of older adult care facilities poses a significant challenge, as no structured data source provides detailed information specifically on these services. Both administrative and crowdsourced data cover general healthcare facilities without indicating whether they specialize in older adult care. To address this gap, a list of keywords was employed to identify healthcare facilities focused on older adults based on their names. However, this method may result in missing relevant facilities, highlighting the need for more precise datasets to better capture the older adult care landscape.

Governmental vs. Non-Governmental Services: Another limitation is the lack of information regarding governmental versus non-governmental services. None of the data sources contain variables to identify this distinction for childcare or older adult care services. To overcome this, a manual search was conducted using keywords such as 'public', 'official', and 'government', which yielded 13 matches for childcare and none for older adult care. Although some public facilities managed by the Ministry of Women and Children Affairs are known, there is no public list with the exact geolocation of these facilities, which limits the analysis.

Childcare Provision in Garment Factories: A specific challenge is the absence of data on childcare provision within garment factories. There is no variable in the data sources to identify such facilities. A Google API search was conducted using the keywords 'Employer-sponsored childcare' and 'Garment factory childcare', but no matches were found. This gap underscores the need for more detailed data to understand the availability of employer-sponsored childcare in industrial settings.

5.2 DEMAND

Outdated data: One of the significant limitations encountered in this analysis is the use of outdated data, particularly concerning slum areas. The most recent data available for slums dates back to 2017, which may not accurately reflect current conditions. Urban landscapes, especially in rapidly developing or changing environments, can undergo significant transformations in just a few years.

Non-disaggregated data at the upazila level: WorldPop uses broad national or city-level distributions to estimate population characteristics. This lack of granularity limits the precision of the analysis for key demographic variables, such as the dependency ratio or the sex ratio, which are crucial for understanding the specific needs and challenges of different communities within upazilas.

Similarly, the INFORM Index provides data at the national level for earthquake risk. The absence of upazila-specific data reduces the ability to conduct a nuanced analysis that takes into account local differences, which are often critical for effective planning and intervention strategies.

Inclusion of Children and Older Adult Individuals with Disabilities: One of the research objectives was to incorporate data on children and older adult individuals with disabilities to provide a more comprehensive analysis. Unfortunately, information on this demographic was not available at the upazila level, resulting in a significant limitation. None of the datasets used for supply mapping include specific variables for disabilities, restricting the ability to assess the availability and adequacy of care facilities for persons with disabilities (PwD). This highlights the urgent need for more inclusive datasets that capture the full spectrum of the population, including PwD. Future research should prioritize the integration of disability variables to address this gap.

5.3 SUPPLY

5.3.1 ADMINISTRATIVE DATA

Classification: For preschool childcare, the dataset includes three types: pre-primary school, kindergarten and nursery, and pre-primary level madrasa. The inability to separate kindergarten and nursery data limits the granularity of the analysis.

In terms of older adult care, the dataset does not specifically identify services tailored to the older adult population. While healthcare facilities are included, there is no variable to distinguish whether they cater to older adults, limiting the accuracy of this analysis.

Lack of Information in Administrative Data: While administrative data provides valuable insights into functional levels and building locations, it suffers from significant issues such as missing values and typographical errors in facility names and addresses. These inaccuracies necessitate careful deduplication efforts. Additionally, the administrative data available for the health sector lacks detailed information specifically related to older adult care, limiting its usefulness for analyzing this particular demographic.

Land use structure: The dataset was originally designed with a focus on land use, presenting several challenges for this study. One significant issue is the presence of duplicate entries for single facilities with multiple buildings. For example, a school with several buildings may appear multiple times in the dataset, necessitating an extensive deduplication process. A distance-based criterion was employed in this project to mitigate this issue. Additionally, many entries lack names, making identification and analysis more complex.

5.3.2 CROWDSOURCED DATA

Coverage: Crowdsourced data supplements the administrative dataset by capturing facilities that might not be included in official records, particularly in education and health, with a focus on kindergartens. However, it excludes other forms of childcare and does not cover older adult care. While this dataset broadens the scope by including lesser-known facilities, it may lack the official validation found in administrative data, potentially affecting its reliability.

Bias: Crowdsourced data is also susceptible to bias, primarily due to the nature of its collection, which relies on voluntary contributions from individuals or organizations. The voluntary aspect introduces several potential biases. First, there may be a geographic bias, where data is more likely to be contributed from areas with higher internet access, digital literacy, and engagement with crowdsourcing platforms. Second, crowdsourced data can reflect the interests and perspectives of the contributors, leading to selection bias.

Additionally, the quality and accuracy of the data can vary widely, depending on the contributors' knowledge and attention to detail. This variability introduces a level of uncertainty into the analysis, as the data may not be uniformly reliable or comprehensive.

5.3.3 GOOGLE API DATA

Misclassification: These irrelevant entries often include parking lots or other non-essential facilities associated with the main identified facility, which can clutter the dataset and obscure meaningful analysis. The API queries take into account not only the names and tags of facilities but also user-generated content, such as reviews. This integration of reviews can lead to the misclassification of some facilities, as user opinions or descriptions may not always accurately reflect the nature or function of the facility.

Lack of granularity: Moreover, the tags provided by the API often lack critical details, making it difficult to distinguish between different types of facilities. For instance, the absence of specific tags or identifiers to differentiate public from private facilities poses a significant challenge. This lack of granularity in the data complicates efforts to categorize and analyze facilities accurately based on ownership or service type. As a result, the data must undergo extensive filtering and manual verification to ensure that only relevant and correctly classified facilities are included in the analysis. These challenges highlight the importance of complementing Google API data with other more structured data sources and suggest that, while powerful, the API should be used cautiously and in conjunction with rigorous data-cleaning processes.

Bias: This data can be subject to several forms of bias. Google's data can be influenced by the nature of user-generated content, which can include inaccuracies, outdated information, or incomplete listings. These biases can skew the data, leading to an overrepresentation of certain types of facilities or regions and an underrepresentation of others, thus affecting the comprehensiveness and reliability of the analysis.

END NOTES

1 Ministry of Local Government, Rural Development and Cooperatives. People's Republic of Bangladesh, 2015

2 See: https://www.nature.com/articles/sdata201789

3 WorldPop, 2020

4 Ministry of Disaster Management and Relief & UN Bangladesh. (2022). Bangladesh INFORM Sub-National Risk Index 2022.

5 Rahaman, S. (2013). Climate Change, Disaster and Gender Vulnerability: A Study on Two Divisions of Bangladesh | American Journal of Human Ecology.

6 Goudet, S. et al. (2010) Impact of flooding on feeding practices of infants and young children in Dhaka, Bangladesh Slums: what are the coping strategies?

7 World Health Organization (WHO). (2023). Saving lives, preventing health risks and restoring health care services in the aftermath of the Libya floods.

8 Gaur, S. & Singh, R. (2023) A Comprehensive Review on Land Use/Land Cover (LULC) Change Modeling for Urban Development: Current Status and Future Prospects

9 World Bank. (2023). Dhaka (Bangladesh) - Land Use / Land Cover Maps (ESA EO4SD-Urban).

10 United Nations Entity for Gender Equality and the Empowerment of Women (UN Women). (2023). Childcare solutions serving low-income communities in Bangladesh. Unpublished manuscript.

11 World Bank. (2023). Dhaka (Bangladesh) - Land Use / Land Cover Maps (ESA EO4SD-Urban).

12 Note the data set for open and green areas extends beyond this category (refer to page 6).

13 United States Environmental Protection Agency (US EPA). (2015). Reduce Urban Heat Island Effect.

14 Huang, S., Gan, Y., Zhang, X., Chen, N., Wang, C., Gu, X., Ma, J., & Niyogi, D. (2023). Urbanization Amplified Asymmetrical Changes of Rainfall and Exacerbated Drought: Analysis Over Five Urban Agglomerations in the Yangtze River Basin, China. Earth's Future, 11(2), e2022EF003117.

15 United Nations Children's Fund (UNICEF). (nd). Extreme heat.

16 World Bank. (2023c). Dhaka (Bangladesh) - Land Use / Land Cover Maps (ESA EO4SD-Urban).

17 In the dataset used for this research, the terms "informal settlements" and "slums" are used interchangeably.

18 Cities Alliance. (2021). Ghost Citizens: women and informality in cities. What can Europe do?

19 European Commission (2022) Bangladesh. INFORM subnational model of Bangladesh.

20 World Bank. (2023a) Dhaka (Bangladesh) - Flood Maps (ESA EO4SD-Urban)

21 Some upazilas have wards under the DNCC and DSCC jurisdiction. When this occurs, the upazila is grouped with the City Corporation that has more wards under its jurisdiction.

22 Hasan, M. et al. (2017). Transforming Urban Dichotomies and Challenges of South Asian Megacities: Rethinking Sustainable Growth of Dhaka, Bangladesh.

DATA-POP ALLIANCE

UN WOMEN