PNAS
www.pnas.org

Supplementary Information for

Global mapping of urban–rural catchment areas reveals unequal access to services

Andrea Cattaneo, Andy Nelson, Theresa McMenomy

Andrea Cattaneo*
Email: andrea.cattaneo@fao.org

This PDF file includes:

Methods
Table S1
Figures: S1-S3
Legend for dataset file

Other supplementary materials for this manuscript include the following:

Dataset S1

Note: Meta-data and geospatial data describing the urban-rural catchment areas, can be found as a GeoTIFF here: https://doi.org/10.6084/m9.figshare.12579572

Supplementary Information Text

Methods

Datasets. For travel time to the nearest city or town, we use global accessibility indicators for the year 2015 following Nelson *et al.* (1).[i] The cities and towns are identified using the Global Human Settlements – Settlement Model grid (GHS-SMOD) (4). To calculate the population, we use the GHS Population grid (GHS-POP) (5) for the year 2015.

Steps for creating the urban-rural catchment areas.

We introduce an algorithm that determines how rural populations gravitate around cities and towns of different sizes. The result is a classification of both urban and rural areas using a consistent definition applied to all countries.

The methodology underlying the calculation is summarized as follows:

1. *Identify and mask out the cities and towns in the GHS-SMOD spatial dataset.* Our cities are the "urban centers" that are coded as typology 30 in the GHS-SMOD. These are high-density clusters of at least 50,000 people. For towns, we use the "urban clusters" that are coded as typology 23.

2. *Calculate the urban population in each city and town agglomeration using the GHS Population grid (GHS-POP) and set population thresholds*. We define seven urban categories: (i) cities that are larger than five million people; (ii) those between one and five million; (iii) 500,000 and one million; (iv) 250,000 and 500,000; (v) 100,000 and 250,000; (vi) 50,000 and 100,000; (vii) towns between 20,000 and 50,000 inhabitants. Note that the GHS typology 23 towns that identified in the previous step have a minimum population of 5,000 people. Therefore, "towns" in our algorithm refers to those urban clusters with more than 20,000 inhabitants. After step 5, the remaining typology 23 clusters with populations between 5,000 and 20,000 inhabitants will be defined as "dispersed towns" if they are located further than 3 hours away from any agglomeration of at least 20,000 people.

3. *Set travel time classifications and create travel masks for the rural pixels around each urban center.* We define peri-urban locations as those facing less than one-hour travel time to the edge of the settlement polygon most accessible to them. Peri-rural locations face between one and three hours travel time to the edge of the most accessible settlement polygon. Using the cost surface accessibility layers, create travel masks around each of the seven urban categories for travel times of less than 1 hour, 1 to 2 hours and 2 to 3 hours. This sets the pixels that fall outside of the travel time category to null and sets 1 to the pixels inside the travel time category.

[i] The global accessibility indicators used here are updated versions of the ones used in Nelson *et al* (1). The update is based on two changes. The accessibility to cities indicators use one different population threshold, 250,000 people here instead of 200,000 people in Nelson *et al* (1). All other population thresholds remain unchanged. The second updated is due to the use of a new cost surface which is the basis for all the accessibility layers. The cost surface represents the time required to cross each pixel of the Earth's surface and is used as an input to the cost accumulation calculation to compute travel time to the nearest city. The cost surface used in Nelson *et al* (1) comes from Weiss *et al* (2018) (2). The updated cost surface used here comes from Weiss *et al* (2020) (3). In all other respects, the methodology to compute the accessibility indicators is identical to that used in Nelson *et al* (1).

4. *Establish a hierarchy to the travel masks based on urban center size and travel time.* Proximity to a larger city "dominates" over a smaller one in the same travel time category and affects the order in which the algorithm is carried out. The blue numbers in Table S1 show the prioritization of urban centers and catchment areas. Urban center 1 dominates over urban center 2, and so on. For the catchment areas, being 1-2 hours from an agglomeration of 50,000 or more (cities) dominates being 0-1 hours from an agglomeration of 20,000-50,000 people (towns). This was decided based on the qualitative difference in services provided once an agglomeration is below 50,000 people. Dispersed towns refers to towns with between 5,000 and 20,000 people that are located over 3 hours travel time to any agglomeration larger than 20,000 inhabitants. Populations are considered to be in the rural hinterland if they are further than three hours away from the edge of an urban center.

5. *Create the urban-rural catchment areas ("catchment masks") by modifying the travel masks according to the hierarchy*. The peri-urban catchment area for the largest city (blue 8 in Table S1) dominates over that of the peri-urban urban center 2 (which ranks 9 in the hierarchy). Where there is overlap between the travel masks of urban centers 1 and 2, pixels in travel mask of urban center 2 are set to null. Next, for step 10 in the hierarchy, pixels in the travel mask of urban center 3 where there is overlap with the travel masks of urban centers 1 and 2 are set to null. Continue to create the catchment areas from the travel masks following the order presented in Table S1. Once all pixels within 3 hours of a city or town of at least 20,000 people have been assigned to a catchment area, identify the "dispersed towns" using the GHSL typology 23 clusters that have populations below 20,000. Lastly, the remaining pixels are the hinterland.

Steps for characterizing the population along the urban-rural continuum by administrative unit.

6. *Identify pixels in each catchment area with high- and low-density populations*. We use a threshold of 1,500 inhabitants per km^2 (pixel) to denote high- and low-density. Using the GHS-POP, create high- and low-density masks. For example, set null to the pixels values below 1,500 and 1 to those above for the threshold for the high-density mask, and the opposite for that of the low-density. Multiply these density masks by the catchment masks to parse out which pixels have high- or low-density.

7. *Calculate the population across the urban-rural catchment areas.* Multiply the catchment masks by the GHS-POP raster.

8. *Tabulate the populations across the urban-rural catchment areas by administrative unit.* We used zonal statistics to sum the populations in each high- and low-density catchment area by country.

Table S1. Hierarchy of dense urban agglomerations by population size and gravitating populations by travel times. The blue numbers represent the order of the urban center and catchment areas in the hierarchy.

		Urban centers	Peri-urban	Peri-rural		Hinterland
			0-1 hours	1-2 hours	2-3 hours	>3 hours to any city
Large cities	> 5 million	1	8	14	21	
	1 – 5 million	2	9	15	22	Dispersed towns 28
Intermediate cities	500,000 – 1 million	3	10	16	23	
	250,000 – 500,000	4	11	17	24	29
Small cities	100,000 – 250,000	5	12	18	25	
	50,000 – 100,000	6	13	19	26	
Towns	20,000 – 50,000	7	20a	20b	27	

Note 1: In the hierarchy of the classification, being 1-2 hours from an agglomeration of 50,000 or more (14 to 19 in the hierarchy shown in blue) dominates being 0-1 hours from an agglomeration of 20,000-50,000 people (20a in the hierarchy shown in blue). This was decided based on the qualitative difference in services provided once an agglomeration is below 50,000 people.
Note 2: Dispersed towns refers to GHSL typology 23 towns (6), but limited to those inhabited by 5,000 to 20,000 people are located over 3 hours travel time to any agglomeration larger than 20,000 inhabitants.

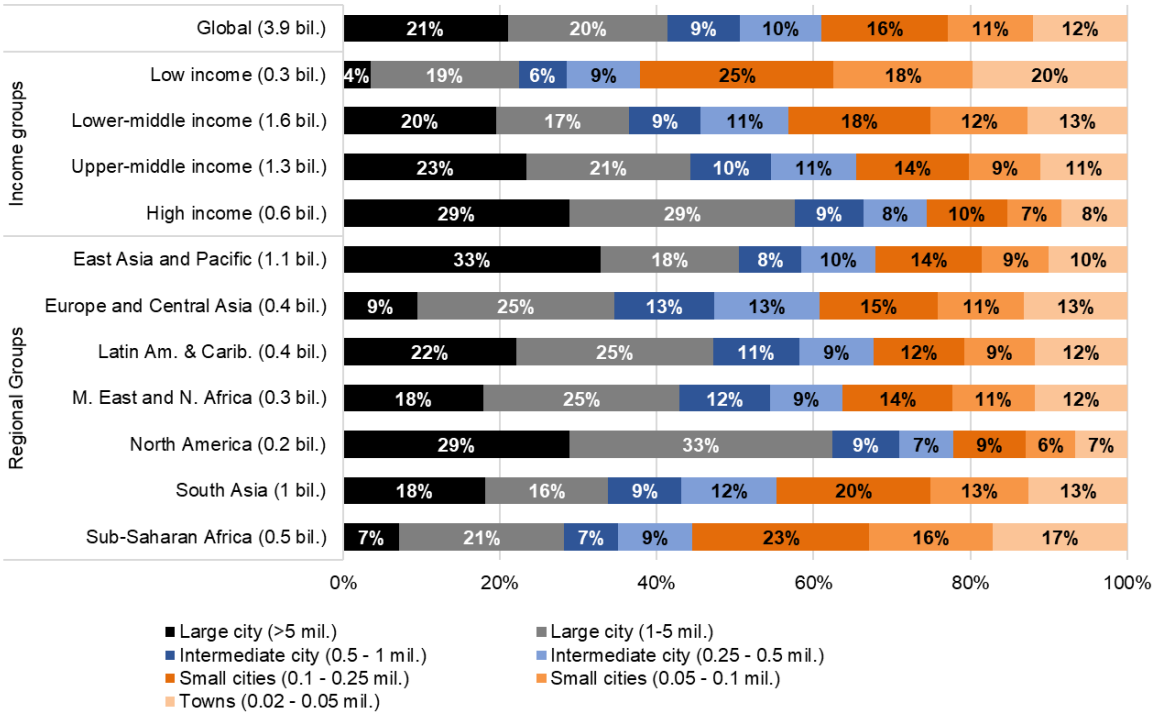

Fig. S1. Global distribution of the urban population in 2015 across cities of different sizes, and by country income and regional groups. Percentage share of the total urban population located in cities of different sizes (with urban population numbers in billions of people shown in parentheses).

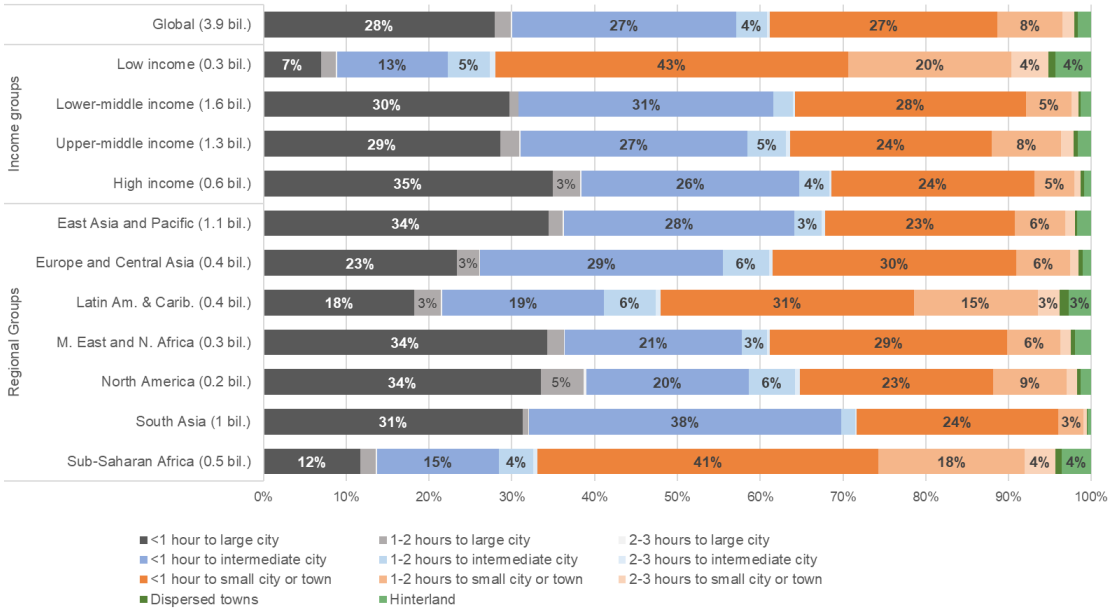

Fig. S2. Global distribution of the rural population in 2015 by travel time to cities of different sizes, and by country income and regional groups. Percentage share of the total rural population (with urban population numbers in billions of people shown in parentheses).

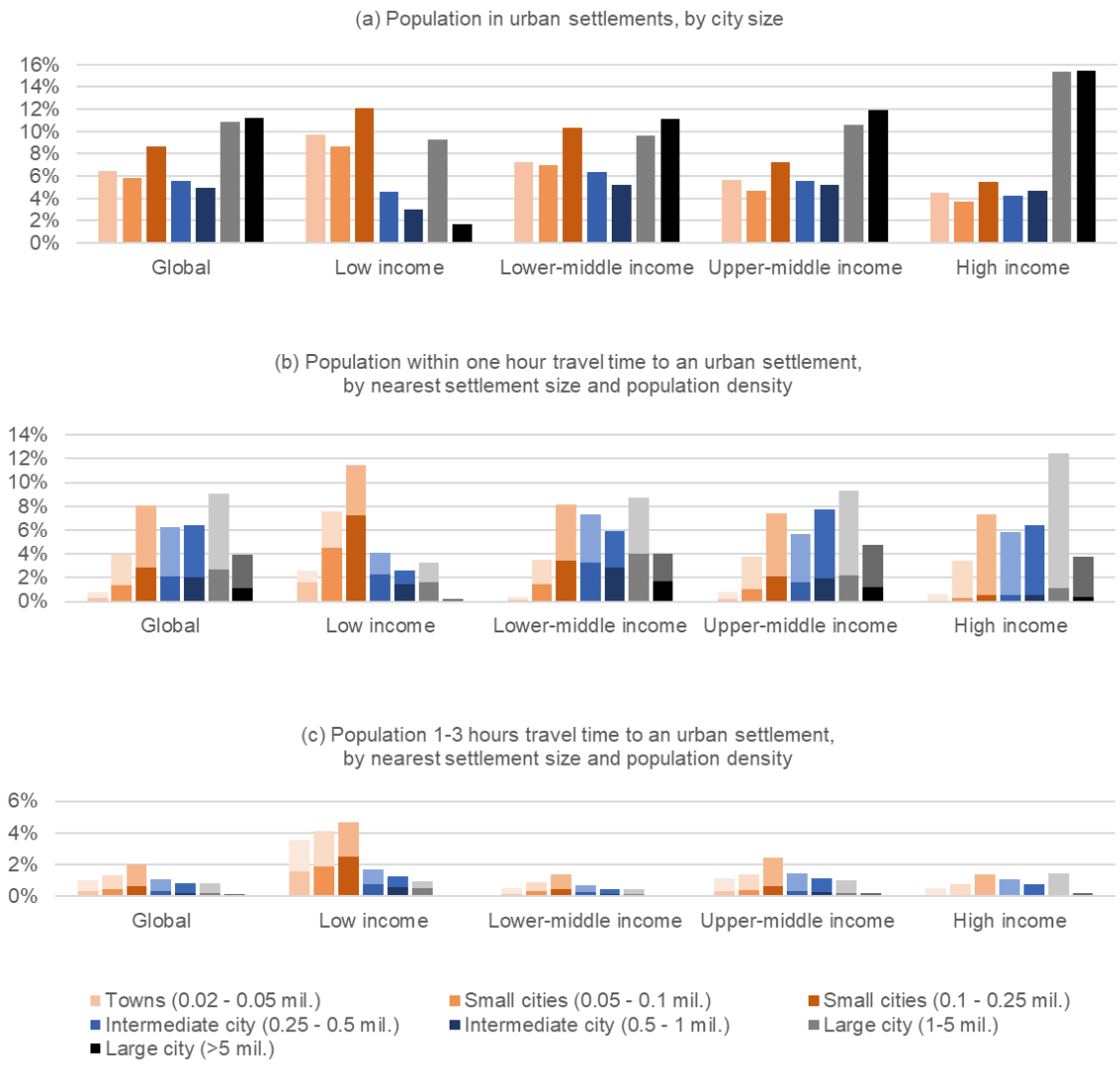

Fig. S3. Global distribution of the population in 2015 at a finer scale, by country income group. Percentage share of the total population located in cities of different sizes or in proximate areas of different minimum travel times to these cities. Lighter shaded insets indicate low-density rural areas (less than 1500 people per km^2); darker insets indicate high-density areas outside of urban agglomerations of 50,000 people or more.

Dataset S1 - Population_URCA (separate file). Population_URCA a tabular dataset containing 2015 population values for the urban–rural continuum at the level of countries and territories. Worksheets (i) to (iv) include the charts and data for Figure 2 in the manuscript and Figures S1 to S4. The ReadMe1 and ReadMe 2 tabs describe the various ways the population data is aggregated in worksheets (v) to (xii). These include the population disaggregated by between those in high- and low-density areas (defined at a threshold of 1,500 people per km^2), and those residing in small rural towns with between 5,000 and 20,000 people. The country-level population values along an urban–rural continuum of 14 categories (aggregated from the 30 categories presented in Table S1), providing the populations of catchment areas for large cities (with populations greater than 1 million), intermediate cities (populations between 250,000 and 1 million), and small cities and towns (between 20,000 and 250,000 people), as shown in Figure 2. The region and country-income grouping for each country is included.

SI References

1. A. Nelson, *et al.*, A suite of global accessibility indicators. *Sci Data* **6**, 266 (2019).

2. D. J. Weiss, *et al.*, A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature* **553**, 333–336 (2018).

3. D. Wiess, *et al.*, Global maps of travel time to healthcare facilities. *Nature Medicine* (accepted).

4. M. Pesaresi, A. Florczyk, M. Schiavina, M. Melchiorri, L. Maffenini, GHS settlement grid, updated and refined REGIO model 2014 in application to GHS-BUILT R2018A and GHS-POP R2019A, multitemporal (1975-1990-2000-2015), R2019A (2019) https:/doi.org/10.2905/42E8BE89-54FF-464E-BE7B-BF9E64DA5218 (June 10, 2020).

5. M. Schiavina, S. Freire, K. MacManus, GHS population grid multitemporal (1975-1990-2000-2015), R2019A (2019) https:/doi.org/10.2905/0C6B9751-A71F-4062-830B-43C9F432370F (August 28, 2020).

6. A. J. Florczyk, *et al.*, “GHSL Data Package 2019” (Publications Office of the European Union, 2019) (July 22, 2020).