

# Differences between gridded population data impact measures of geographic access to healthcare in sub-Saharan Africa

**Fleur Hierink** (✉ [fleur.hierink@unige.ch](mailto:fleur.hierink@unige.ch))

University of Geneva <https://orcid.org/0000-0002-2727-0540>

**Gianluca Boo**

University of Southampton

**Peter Macharia**

KEMRI Wellcome Trust Research Programme <https://orcid.org/0000-0003-3410-1881>

**Paul Ouma**

Kenya Medical Research Institute - Wellcome Trust Research Programme

**Pablo Timoner**

GeoHealth, University of Geneva

**Marc Levy**

Columbia University <https://orcid.org/0000-0002-1111-2222>

**Kevin Tschirhart**

Earth Institute at Columbia University

**Stefan Leyk**

University of Colorado <https://orcid.org/0000-0001-9180-4853>

**Nicholas Oliphant**

The Global Fund to Fight AIDS, Tuberculosis, and Malaria

**Andrew Tatem**

University of Southampton <https://orcid.org/0000-0002-7270-941X>

**Nicolas Ray**

University of Geneva <https://orcid.org/0000-0002-4696-5313>

---

## Article

## Keywords:

**Posted Date:** April 28th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1514074/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Communications Medicine on September 16th, 2022. See the published version at <https://doi.org/10.1038/s43856-022-00179-4>.

# Abstract

Access to health care is imperative to health equity and well-being. Geographic access to health care can be modelled by combining different spatial datasets, among others, on the distribution of existing health facilities and populations. Several population datasets are currently available, but their impact on accessibility analyses is unknown. In this study, we model the geographic accessibility of public health facilities at 100-meter resolution in sub-Saharan Africa and explore the effect of six among the most popular gridded population datasets on coverage statistics at different administrative levels. We found differences in accessibility coverage of more than 70% at the sub-national level, based on a one-hour travel time threshold. Differences are significant in large and sparsely populated administrative units, dramatically shaping patterns of health care accessibility at the national and sub-national level. The results underscore an essential source of uncertainty in accessibility analyses that should be systematically assessed in policy-making.

## Introduction

Geographic access to healthcare is essential to ensure universal health coverage, a key target of the United Nations Sustainable Development Goals (SDGs)<sup>1</sup>. While geographic access is only one of many factors, such as affordability, availability, and acceptability<sup>2,3,4</sup>, that impacts access to healthcare, it is fundamental to the organization of a health system<sup>5,6</sup>. Modelling geographic access to healthcare is necessary to identify gaps in health system coverage and to support targeted health system optimization and planning, such as placement of new facilities, deployment of community health workers, or mobile outreach<sup>7</sup>. The key components of a geographic accessibility analysis are the population needing access, the locations of health facilities, and data to help model connectivity and travel time (i.e., road networks, landcover, streams, elevation and care seeking specificities)<sup>5,8</sup>. Although data on each of these components is increasingly available, accurate and current<sup>9</sup>, there are persistent differences between regions, hampering accessibility analyses in data-poor regions<sup>10</sup>. Global advancements in population modelling have enabled the research community to use several gridded population datasets<sup>11,12,13,14,15,16</sup>, in combination with recent data on health facility location<sup>17</sup>, opening new avenues for modeling geographic accessibility to healthcare in data-poor settings. The effects of using different population data in accessibility analyses to calculate the proportion of the population that can access a health facility within a given travel time threshold, i.e., *accessibility coverage*, and thus monitoring indicators that underpin policy-making at the global, national, and sub-national level, are unknown. This study aims to shed light on the magnitude and variation of these effects and possible policy implications, by conducting the first comprehensive comparison of six of the most commonly used global gridded population datasets in a geographic accessibility model at 100-meter resolution for sub-Saharan Africa.

Gridded population datasets allocate population counts across rows and columns of grid cells either by using simple techniques to uniformly redistribute census data or by using Earth observations, population data, and dasymetric modelling techniques, that provide more refined estimates<sup>18</sup>. These datasets

typically use a country's most recent census or projected estimates, summarized in available administrative units or census enumeration areas, to disaggregate population numbers at a higher spatial and temporal resolution<sup>19,20</sup>. Population redistribution techniques vary from dataset to dataset, meaning that the suitability of each dataset for any spatial analysis is context-dependent. Discrepancies between datasets do not necessarily reflect specific appropriateness; rather the suitability of each gridded population dataset is highly dependent on the target scale, context and purpose, and geographic extent of the analysis<sup>18</sup>. However, even when two or more gridded population datasets meet some predetermined criteria, differences in accessibility coverage may be observed. Different population data have been used in accessibility analyses, exposing potential uncertainty in accessibility coverage estimates and making comparability across studies difficult. Some studies have used national censuses<sup>21</sup>, WorldPop products<sup>7, 10, 22, 23, 24, 25, 26</sup>, Gridded Population of the World (GPWv4)<sup>27</sup>, High Resolution Settlement Layer (HRSL)<sup>28</sup>, or Landscan<sup>29</sup>. The scientific literature increasingly acknowledges differences between gridded population datasets<sup>18,19</sup>. However, often the focus is on general data properties and assessment of appropriateness<sup>18,19,30</sup> or country- or discipline-specific implications of using the various data products<sup>20,31,32,33</sup>. In addition, the motivation and implications of using a particular population dataset are usually neglected in accessibility studies<sup>33,34</sup>. The choice of any specific population layer is likely driven by user personal preferences, lack of knowledge of other sources, or ease of access and use.

Here, we systematically assess differences between estimates of geographic healthcare accessibility for all of sub-Saharan Africa using the most popular gridded population data products: 1) WorldPop top-down constrained, 2) WorldPop top-down unconstrained, 3) HRSL, 4) GPWv4, 5) Landscan, and 6) Global Human Settlement Population (GHS-POP). Healthcare accessibility is modelled at 100-meter resolution using the most recent release of the geo-coded health facility inventory of 50 countries in sub-Saharan Africa to enable a fair comparison across models<sup>17</sup>. We contrast accessibility coverage statistics derived from the six population datasets (Supplementary Table 1), across countries at national and sub-national scale. Travel time to the nearest health facility was calculated by developing a friction layer representing the estimated time required to reach the nearest health facility. We intersected various gridded population datasets with our travel time estimates to determine accessibility coverages within various travel time thresholds (i.e., 30, 60, 90, 120, 150, and 180-minutes). Our accessibility coverages vary widely between the different datasets and estimates on the sub-Saharan African level mask larger sub-national variations. Differences are most pronounced in scarcely settled regions, where administrative units are large. Datasets that distribute population over larger land areas rather than limiting to building footprints can drastically change accessibility patterns. The results provide useful clues for policy-making and critically reflecting on earlier estimates of accessibility to health care and their associated uncertainties.

## Results

### *Accessibility coverage estimates*

#### *Sub-Saharan Africa*

Estimates of accessibility coverage, modelled by constructing a travel time grid at 100-meter resolution for all of sub-Saharan Africa (Supplementary Fig. 1), show greatly divergent results using the six different population datasets (Fig. 1A and B). For all of sub-Saharan Africa, the population that has access to healthcare is highest when using HRSL, followed by GHS-POP (Fig. 1B). Differences in accessibility coverage are larger at 30- and 60-minute catchments and logically decrease as travel times increase. An estimated 88.2% of the HRSL-derived population has access to a health facility within 30 minutes travel time. This value drops to 60.5% when GPWv4 is considered. Access to healthcare is in general substantially lower when statistics are derived using GPWv4 and WorldPop top-down unconstrained datasets (Fig. 1B). These two datasets also present the largest differences in accessibility coverage as compared to the other datasets (Fig. 2). Although the differences between the other datasets are smaller, there are still coverage differences of up to 9.5% among the other population products at 30 minutes travel time (Fig. 2). The relative differences are smallest between Landscan and WorldPop top-down constrained and between HRSL and GPWv4. Accessibility coverages at the sub-Saharan African level already show strong variation, but continental summary statistics substantially mask variations at the national and sub-national level.

### ***National***

Moving from sub-Saharan African to national coverage statistics, we find new patterns, with varying results between and within countries (Fig. 3 and Supplementary Table 2). Strongly divergent trends are particularly evident in some countries, including Chad, Sudan, Eritrea, South Sudan, Central African Republic, Republic of the Congo, Democratic Republic of the Congo, Equatorial Guinea, and Gabon (Fig. 3). In these countries, we observe lower coverage statistics for GPWv4 and WorldPop top-down unconstrained, sometimes followed by significant discrepancies between the coverage values for the other datasets. The differences in accessibility coverage can exceed 60% and would affect any conclusion drawn from one of the individual population datasets. In the Republic of the Congo, for example, accessibility coverage at 30 minutes travel time ranges from 28.8% to 88.9%. Using GPWv4 or WorldPop top-down unconstrained suggests that 71.2% or 65.5% of the population in the country is unable to reach the nearest health facility within half an hour travel time. In contrast, using GHS-POP, HRSL, Landscan, or WorldPop top-down constrained indicates that 11.1%, 13.9%, 15.8%, or 27.3% of the population is unable to reach healthcare within half an hour. This discrepancy between the datasets may have a strong impact on the conclusions drawn from monitoring global and national indicators of access to healthcare, and thus on decision making for resource allocation.

### ***Sub-national***

Figure 4 illustrates accessibility coverage within 1-hour catchments at the sub-national (i.e., administrative 1) level. Supplementary data 1 presents accessibility coverage for 30, 60, 90, 120, 150, and 180 minutes travel time at administrative level 2. Despite the similarities in overall accessibility patterns, with low access in northern and central sub-Saharan Africa and higher access in southern sub-Saharan Africa and coastal regions, sub-national differences between the datasets are clearly evident. Low

accessibility coverage is particularly widely spread for GPWv4 and WorldPop unconstrained. In Figure 5 we present the average percentage point difference between the datasets we observe at the sub-national level. The average difference between all datasets can be as high as 45.4%. However, when comparing individual datasets, the sub-national average difference can exceed 70% (Fig. 5B).

### ***Explaining discrepancies in coverage estimates***

Most of the observed discrepancies in accessibility coverage can be explained by the characteristics and quality of the input data and the redistribution approach used for creating the gridded population datasets. More specifically, the main differences in accessibility coverage that we observe can be explained by 1) the use of settlement data to conditionally constrain population to buildings, 2) the quality and resolution of the settlement data used, and 3) the granularity of the smallest publicly available unit for population data. In Figures 4 and 5, the differences in accessibility coverage are particularly evident between datasets that constrain population to settlements (i.e., WorldPop top-down constrained, HRSL, GHS-POP, and Landscan) and the other datasets that allocate population based on dasymetric weighting or other areal interpolation techniques. Constrained population datasets typically use building footprints or settlement feature data derived from satellite imagery to constrain the distribution of population to grid cells in which buildings have been detected. The datasets based on settlement data have a large proportion of zero cells in areas where no buildings are detected<sup>31</sup>. This means that population is commonly distributed over smaller areas and therefore more concentrated in regions with human activity and health facilities. In contrast, datasets that do not contain information on settlements have a small proportion of zero cells. This is a natural consequence of using approaches that spread population over vast areas of land where few or no people are likely to reside, including extremely uninhabitable areas such as deserts or dense forests where there are no health facilities. These distorted distributions ultimately result in longer travel times for some of the population and therefore smaller overall accessibility estimates.

In northern Chad, for example, accessibility coverage is between 58.1% and 72.4% at 30 minutes travel time using HRSL, GHS-POP, Landscan, or WorldPop top-down constrained, and drops to almost 0% when GPWv4 or WorldPop top-down unconstrained is considered. Similar patterns were also observed in northern Niger and other regions south of the Sahara desert. This region is sparsely populated and has large differences in accessibility patterns between the datasets. Figure 6 shows an example of the observed visual differences between the datasets. The same is true for some regions in central sub-Saharan Africa, such as the Republic of the Congo, Gabon, and the Democratic Republic of the Congo where large areas of land are characterized by dense and closed forests with very few detected settlements (Supplementary Fig. 2). In Ogooué-Maritime, a province in western Gabon characterized by dense forests, accessibility coverage within 30 minutes ranged from 87.9% to 96.3% when using WorldPop top-down constrained, Landscan, HRSL, or GHS-POP, in ascending order of coverage. However, accessibility coverage decreases to 11.1% and 3.8% when WorldPop top-down unconstrained and GPWv4 are used. Comparisons of accessibility coverage between the settlement-based population data also show discrepancies (Fig. 2 and Fig. 5), as their accuracy appears to be highly dependent on the complete

identification of individual building structures. The quality of the underlying satellite data containing information on built environments and the applied methodology to automatically extract built features involves omission and commission errors, leading to an under- or overestimation of uninhabited areas<sup>20, 35, 36</sup>. While WorldPop top-down constrained uses polygon building footprint data and HRSL uses high resolution satellite imagery (~50 cm), GHS-POP extracts built features from Landsat 8 imagery with a resolution of ~30 meters<sup>30</sup>. Due to the difficulty of interpreting built-up areas from coarser satellite imagery, GHS-POP and, to a lower extent, Landscan have previously been found to overestimate uninhabited zones and thus underestimate people in sparsely populated sub-urban and rural areas<sup>20, 37, 38</sup>. We found similar patterns in two rural areas in Garissa and Nakuru counties in Kenya, where divergent patterns of settlement detection between the gridded population products were seen (Supplementary Figures 3 and 4). Particularly GHS-POP did not seem to allocate population in small settlements that were included in the other datasets (Supplementary Figure 4). When no population is allocated to small rural settlements, a relatively large proportion of the population is distributed into larger built areas where facilities are located, this likely contributes to higher accessibility coverage statistics for GHS-POP and Landscan as compared to HRSL and WorldPop top-down constrained.

An important challenge for all gridded population datasets is the quality and granularity of the input population data. Even though census data is often collected at the household level or in smaller enumeration areas, countries usually release aggregated data at specific administrative levels to protect privacy<sup>18</sup>. The scale at which the latest population census is made publicly available varies widely across sub-Saharan Africa (Figure 7A) and ranges on average from about 2 km<sup>2</sup> to 182,211 km<sup>2</sup>. Figure 7B illustrates the association between population input unit size (km<sup>2</sup>), relative coverage difference between the datasets at 1-hour travel time, and average total population per administrative unit (level 1). The figure shows that in areas where there are large differences in accessibility coverage between the datasets, the size of the population input unit is generally large, and the total population living in these units is small, mostly in the first or second quantile (Fig. 7B, top right corner). This means that when population counts in thinly populated areas are aggregated into large units, differences between the datasets are greatest. Figure 5 and Figure 7 show similarities between areas with high accessibility coverage differences and regions with large population input sizes, such as the northern- and central parts of sub-Saharan Africa. Sangha, for instance, a region in the Republic of the Congo has one of the highest average accessibility coverage differences between all datasets (45.4%). The average total population of 45,281 people is spread out over approximately 57,686 km<sup>2</sup> land and the landscape primarily exists of dense forests, complicating building detection. The same is true for an area that we described before, Ougououé-Maritime province in Gabon, where the average coverage difference is 45%, the average total population is 44,230, the population input unit size is 7,528 km<sup>2</sup>, and the landscape is dominated by dense forests (Supplementary Figure 2).

The aggregated nature of the input population data masks the spatial variability in population distribution at finer scales and therefore causes uncertainty when total population counts are reallocated into grid cells. Our analysis suggests that particularly in thinly populated areas where population data is

made available at a coarse scale, the different distribution techniques between the datasets cause the most varying reallocation patterns and thus translate into widely ranging accessibility coverage estimates.

## Discussion

Data on population distribution is the main denominator for almost all public health interventions. The effectiveness of evidence-based health planning, such as the distribution of health facilities or the implementation of vaccination campaigns, largely depend on accurate population estimates<sup>39, 40</sup> to calculate resource needs and measure the impact of interventions<sup>41, 42</sup>. Moreover, the SDGs and other international health targets are based on indicators that reflect the proportion of the population that has access to certain services. Knowing how many people live where is essential for these calculations<sup>41</sup>. Here we show that estimates of healthcare coverage vary widely depending on the gridded population dataset chosen and that they can lead to conflicting conclusions.

Our results show notable variations and tend to diverge most in regions with a low population density where administrative units are large, and land cover classes such as dense forests and deserts indicate scarce population distribution. The large variability in our results would also hold true for coverage estimates of other types of services for which similar accessibility models can be used, such as school access for children within a predefined threshold or for estimates related to the people at risk of infectious diseases<sup>43, 44</sup>, people living in disaster prone areas<sup>31</sup>, or modelling vaccination coverage<sup>45, 46</sup>.

The use of one population dataset can have strong implications for policy- and decision-making. With new global targets aimed at improving access to healthcare it is crucial that indicators that monitor progress are correct and based on realistic input parameter values. For instance, the recently adopted target 4 indicator of the World Health Organization (WHO)'s *Ending preventable maternal mortality* (EPMM) strategies states that by 2025 (1) at least 60% of the global population should be able to access the closest functional emergency obstetric care (EmOC) health facility within two hours travel time, and (2) 80% of countries should have a 2-hour accessibility coverage greater than 50%<sup>47</sup>. A United Nations (UN)-led guidance to help countries model this indicator will be released in 2022. In that context, our results can provide useful quantification of the expected relative differences for and thus the sensitivity of this indicator based on various population datasets. Taking again the Republic of the Congo as an example, we find that when using different population datasets, accessibility coverage at 2-hour travel time ranges from 44.7% to 95.0%. Coverage statistics are highest when using GHS-POP (95.0%), HRSL (93.2%), Landscan (93.4%), or WorldPop top-down constrained (84.3%). However, when using WorldPop top-down unconstrained or GPWv4, coverage was considerably lower at 51.5% and 44.7% respectively. This means that our observed differences could lead to very different conclusions when considering thresholds for accessibility coverages, such as EPMM. Supplementary Figure 5 shows the sub-national discrepancies in accessibility coverage at a 2-hour travel time threshold at administrative

level 1. Although the differences are smaller than those in Figure 4 (1-hour travel time), the unconstrained datasets show markedly different patterns than the constrained datasets.

In light of previous research and policy documents, that have relied on a single gridded population dataset for coverage estimates, our results also provide interesting clues for comparison for the same region or country. For example, studies on geographic access to care in Mozambique have used GPW<sup>27</sup>, WorldPop top-down unconstrained<sup>48</sup>, and HRSL<sup>28</sup>, leading to different estimates. In addition, Weiss et al.<sup>34</sup> reported an accessibility coverage of 78.7% in a 1-hour catchment in Madagascar, yet our coverage estimates range from 58.7-76.6% in the same catchment. The same holds true for other countries where estimates from Weiss et al.<sup>34</sup> were either outside our range of estimates or showed large intervals around the reported estimate. While any comparison of coverage is also influenced by other input data used in an accessibility analysis, such as travel scenarios, road networks, and health facility coordinates, our results can be used to get a sense of the potential uncertainty in the coverage estimate as a result of the population denominator chosen.

Gridded population datasets heavily rely on the recency and quality of population censuses, which countries commonly conduct every 10 years, however many countries in sub-Saharan Africa have not conducted a full population census in more than 15 years due to financial constraints, political instability, or remoteness<sup>18, 49</sup>. In the Democratic Republic of the Congo, the last complete census was carried out in 1984 and policy-makers and gridded population data providers therefore rely on imprecise estimates of current population through linear population projections<sup>50</sup>. The growing number of bottom-up population estimation approaches overcomes this challenge by conducting micro-censuses in small areas which are then extrapolated to larger administrative units using ancillary satellite data<sup>41</sup>. In January 2022, WorldPop released bottom-up population estimates for seven provinces (i.e., Haut-Katanga, Haut-Lomami, Ituri, Kasai, Kasai Oriental, Lomami and Sud-Kivu) in the Democratic Republic of the Congo<sup>51, 52</sup>. Interestingly, comparing the relative coverage estimates of the bottom-up and top-down datasets in these seven provinces did not lead to different patterns than earlier observed (Supplementary Table 3), meaning that the relative bottom-up coverage fell within the range of the constrained top-down datasets. However, absolute comparisons were markedly different, with generally lower total population counts in the bottom-up dataset and thus proportionately lower numbers of people falling within the 1-hour health facility catchment. Even though it is impossible to indicate the best gridded population dataset, objective comparisons of population products can improve our understanding of the differences and the implications of using one dataset in particular<sup>19</sup>.

In terms of fitness for use, population datasets that constrain population, based on high resolution settlement data (i.e., HRSL, WorldPop top-down constrained), to settlements are more suited for accessibility modelling assuming acceptable levels of accuracy<sup>42</sup>. Most accessibility models need to consider the population at their place of residence (i.e., *de jure/de facto* population)<sup>19</sup>, because the aim is to capture the complexity of the patient's journey to reach a health facility from their home, so that health system improvements can be targeted, and microplanning of outreach is possible<sup>39, 40</sup>. This is

complicated when datasets do not constrain population to buildings or when ambient population is modelled and thus make GPWv4, WorldPop top-down unconstrained, and Landscan less favorable. However, the interpretation of built-up areas from satellite imagery is not without error. This means that in the absence of complete settlement data, these unconstrained datasets are still important and useful in ensuring that no population is overlooked in health estimates. Resolution and recency are other important factors that weigh in this decision. Accessibility models are dependent on a more local target scale for analysis, the unconstrained population datasets cannot provide this analytical scale because the population is distributed over larger units. Constrained models focus more explicitly on the units of interest relevant for accessibility (areas) and therefore the estimates are more plausible. The HRSL and WorldPop top-down constrained datasets seem the best fit for use when it comes to accessibility modelling to healthcare. Our advice would be to, where possible, consider both datasets and construct a plausible range of coverage statistics, comprising the mean, and a lower and upper bound around the statistics.

There are a number of caveats related to our accessibility analysis which are explained in more detail in the Methods. In brief, our travel time raster was developed while making large scale assumptions on the speeds- and modes of transports which may not appropriately reflect local contexts. We think that by keeping the speeds and modes constant per country, we were able to more concretely discuss the differences in coverage as a result of the population datasets. In addition, we did not reallocate population falling on barriers (i.e., rivers and lakes), however percentages are overall small and not expected to change any of the patterns we observed.

Despite an increasing global availability of data on numerous geographic objects, we still face challenges in precisely and accurately locating population, especially in low- and middle-income countries and areas sparsely populated. Yet, data on population density and distribution are vital inputs for research and policy-making<sup>53</sup>. The results presented in this study show how valuable and critical a comparative analysis between population datasets is for the derivation of coverage statistics that inform local policies and monitor global targets. Our results show that large differences exist between the datasets. This is evenly true for datasets informed by building footprints, even though to a smaller extent. Caution should be taken when drawing conclusions from any single gridded population dataset and potential uncertainties and limitations should increasingly be acknowledged in accessibility studies. A critical comparison of the results provided here shines a light on the reliability and plausibility of coverage statistics.

## Methods

### **Accessibility model:**

Accessibility to healthcare was modelled in terms of travel time to the nearest public health facility. This calculation was made by overlaying health facility coordinates on top of a friction raster. Each grid cell in the friction raster represented a unique land cover class which was assigned a travel speed. On-road

travel represented motorized speeds whereas for off-road travel walking speeds were used. The accumulated time it takes to traverse all cells to the nearest health facility was then calculated for each grid cell, representing the travel time raster. This calculation was done on the eight-directional least-cost path algorithm<sup>8, 54</sup> and was isotropic, meaning that no corrections were made for slopes.

The friction raster represents information about potential impacts on a patient’s journey to healthcare, including land cover type, barriers to movement, and the road network. All this information was extracted from open data sources (Table 1). We fully automated the entire workflow in an R and Python environment. In brief, road networks, rivers, and lakes were extracted from OpenStreetMap (OSM) using the *osmextract*<sup>55</sup> library in R<sup>56</sup> (version 4.0.4). The land cover for sub-Saharan Africa was downloaded at 100-meter resolution from Copernicus<sup>57</sup>. Health facility coordinates were extracted from a geocoded database for sub-Saharan Africa<sup>17</sup>. Administrative boundaries for all African countries were taken from the database of Global Administrative Areas (GADM)<sup>58</sup>.

**Table 1**  
**Overview of data sources for accessibility analysis.**

| Dataset                          | Producer                           | Resolution  | Year | Citation |
|----------------------------------|------------------------------------|-------------|------|----------|
| Landcover                        | Copernicus                         | ~100 meters | 2019 | 57       |
| Roads                            | OpenStreetMap                      | vectorized  | 2021 | 55, 59   |
| Waterbodies (lines and polygons) | OpenStreetMap                      | vectorized  | 2021 | 55, 59   |
| Health facilities                | Maina et al. (2019)                | vectorized  | 2018 | 17       |
| Travel scenario                  | Adapted from Weiss et al. (2020)   | -           | -    | 34       |
| Administrative boundaries        | Global Administrative Areas (GADM) | vectorized  | 2020 | 58       |

Data preparation was done on a per-country basis and optimized to minimize computation time as detailed in the Supplementary Figure 6, implying that land cover data was first downloaded for the entire African continent and then processed for each country, separately. In summary, and as shown in Supplementary Figure 6, data processing included cropping to the bounding box of each country to minimize computation time in the masking step. Then rasters were clipped to exact country borders. Lastly, the land cover raster was projected in the country’s coordinate system (Supplementary Table 4).

The process was parallelized using the *doParallel*<sup>60</sup> and *foreach*<sup>61</sup> R libraries. All necessary data processing steps were done using the *terra* package<sup>62</sup>. Scripts for data processing and analysis can be sourced from Zenodo [will be created after finalization of article].

Vector data representing road networks and barriers to movement were fetched using the *osmextract*<sup>55</sup> library in R<sup>56</sup> (version 4.0.4) and projected in the country's coordinate system (Supplementary Table 4). All road classes that are officially classified by OSM were included for analysis<sup>63</sup>. Barriers to movement (unless a road crosses over) included hydrographic lines classified as river and hydrographic polygons. Streams and smaller waterbodies were excluded from the analysis since they can be traversed with ease<sup>63</sup>.

The geocoded inventory of public health facilities in Sub-Saharan Africa<sup>17</sup> assembled between 2012 and 2018 was downloaded and projected to match the spatial coordinate system of the other datasets by country (Supplementary Table 4). We included all health facilities irrespective of type (e.g., primary, secondary, health centers, etc.).

Finally, all data were combined in a friction raster at 100-meter resolution. This resolution offered the best compromise between computational efficiency, spatial detail to address fine-scale disparities in healthcare access, and consistency with the assembled spatial data described above. The vector data were rasterized at 100-meter resolution. All raster cells were aligned, and layers merged to create one comprehensive land cover raster, to which travel scenarios (Supplementary data 2) were applied. The travel scenarios for all sub-Saharan African countries were taken from Weiss et al. (2020)<sup>34</sup> but adapted to the context of this paper (Supplementary data 2). When a travel scenario from Weiss et al. (2020)<sup>34</sup> did not indicate a speed for a specific road class in a given country, we used the African average travel speed for that road class (Supplementary data 2).

### **Data processing of population grids:**

Population rasters were clipped to country borders and reprojected to each country's projection system (Supplementary Table 4). Population that was lost from the original files, due to these data processing steps, were equally smoothed out over the rasters so that total population counts remained the same as in the original files. This was done by comparing the summed population at administrative level 2 for the original and projected rasters. Due to the different resolutions of the datasets, and to avoid resampling population raster data, all population grids were transformed into spatial points representing the centroids of the grid cells.

### **Extraction of accessibility coverage statistics:**

To assess the spatial variation in national and sub-national accessibility coverage statistics, we overlaid the six gridded population datasets onto the travel time rasters for each country. We extracted the travel time and the administrative boundary (level 1 & 2) for each population point feature. We then calculated the accessibility coverage statistics, by means of zonal statistics, to output the population able to reach the nearest health facility within a certain travel time. Both relative and absolute coverage statistics were obtained per administrative unit (level 1 and national). Population falling on barriers (i.e., water bodies or

just outside country borders) were not included in the extraction of coverage statistics. The absolute and relative number of people falling on barriers is indicated in the Supplementary data 2.

### **Limitations of method:**

We note that to derive the presented coverage statistics, our travel time raster capturing accessibility to the nearest health facility served as the main input data. However, we recognize that realistic estimates of geographic access to healthcare require local knowledge of healthcare seeking behaviors, such as travel modes and speeds, and information about (seasonal) barriers to mobility. Although we have used local expert knowledge to inform accessibility models in previous studies<sup>26, 28, 64</sup>, the scale and context of the presented analysis did not allow us to use such local knowledge. Such detailed input was beyond the scope of this study that is intended to reflect important differences between the population datasets. Thus, while our travel time maps and associated accessibility coverages estimates should not be used for health system planning at national and sub-national scales, our methodology can be adapted to the local context, informed by expertise from various stakeholders at the national and sub-national level, notably on modes and speeds of transport.

One limitation of the current study is that the unconstrained datasets had a proportionately higher number of people falling on areas considered as barriers (i.e., water bodies or areas falling outside of country borders). On the full sub-Saharan African scale this proportion was only 0-2% of the total population, however nationally and sub-nationally these numbers were higher, especially in small island nations.

In addition, modelling geographic accessibility presents challenges other than differences between gridded population datasets. For example, uncertainties in travel modes and speeds can lead to under- or overestimation of accessibility. If travel speeds are assumed to be higher than they actually are, the accessibility model results will incorrectly indicate a higher accessibility coverage. This also applies to uncertainties in road network data when some roads maybe missing or when roads may actually present dirt tracks that in reality cannot be traveled by motorized vehicles. Realistic modelling of access to healthcare therefore depends heavily on reliable and locally agreed-upon model inputs. One nascent area is the use of Google map APIs to characterize travel time that has been shown to estimate near to reality travel times in urban areas. The approach potentially accounts for traffic, weather conditions, difference in speeds, road conditions and other predisposing factors. However, the approach is still at its early stages of development and more applicable in urban areas where through volunteered geographic science, the coverage of data is better relative to remote and rural areas where the majority of the people live. Therefore, the use of least-cost path algorithms still remains feasible but requires improved parameterization<sup>65</sup>.

## **Declarations**

### **Acknowledgements**

PMM is supported by the Royal Society Newton International Fellowship (NIF/R1/201418) and is also grateful for the support of the Wellcome Trust to the Kenya Major Overseas Programme (# 203077). This research benefited from support through the National Science Foundation (award #1924670), and support provided to the University of Colorado Population Center (CUPC, Project 2P2CHD066613-06) from the Eunice Kennedy Shriver Institute of Child Health Human and Human Development. The content is solely the responsibility of the authors and does not necessarily represent the official views of NIH or CUPC.

## Author contributions

FH led the study along with NR. Conceptualization of the study was done by FH, NR, and GB. The methodology was initially developed by FH, NR, and GB, and reviewed by all co-authors. Data analysis and processing was done by FH and supported by PT, GB and NR. Writing of the original draft was done by FH and supported by NR and GB. Initial reviews on the figures were given by PMM, PO, GB, AJT and NR. The manuscript was edited by FH, GB, PMM, PO, PT, KT, NO, ML, SL, AJT, and NR. All authors have further assisted in thoroughly reviewing all figures and texts.

## Data availability

All open data used in this study is available via the relevant resources as indicated in Table 1 and Supplementary Table 1. The code for data processing and analyses are available upon request. All results that were obtained are also accessible in Supplementary data 2.

## Competing interests

The authors have no conflicts of interest to declare.

## References

1. United Nations. Sustainable Development Goal 3: Ensure healthy lives and promote well-being for all at all ages. (2016).
2. Levesque Jean-Frederic, Harris Mark F, Russell Grant. Patient-centred access to health care: conceptualising access at the interface of health systems and populations. *International journal for equity in health* **12**, 18 (2013).
3. Peters David H, Garg Anu, Bloom Gerry, Walker Damian G, Brieger William R, Hafizur Rahman M. Poverty and access to health care in developing countries. *Annals of the New York Academy of Sciences* **1136**, 161-171 (2008).
4. Gulliford Martin, *et al.* What does 'access to health care' mean? *Journal of health services research & policy* **7**, 186-188 (2002).
5. Delamater Paul L, Messina Joseph P, Shortridge Ashton M, Grady Sue C. Measuring geographic access to health care: raster and network-based methods. *International journal of health geographics*

- 11, 15 (2012).
6. World Health Organization. Primary health care measurement framework and indicators: monitoring health systems through a primary health care lens. *Geneva: World Health Organization and the United Nations Children's Fund (UNICEF)*, (2022).
  7. Ouma Paul O, *et al.* Access to emergency hospital care provided by the public sector in sub-Saharan Africa in 2015: a geocoded inventory and spatial analysis. *The Lancet Global Health* **6**, e342-e350 (2018).
  8. Ray Nicolas, Ebener Steeve. AccessMod 3.0: computing geographic coverage and accessibility to health care services using anisotropic movement of patients. *International journal of health geographics* **7**, 63 (2008).
  9. Forget Yann, Linard Catherine, Gilbert Marius. Supervised classification of built-up areas in sub-Saharan African cities using Landsat imagery and OpenStreetMap. *Remote Sensing* **10**, 1145 (2018).
  10. Pu Qiang, Yoo Eun-Hye, Rothstein David H, Cairo Sarah, Malemo Luc. Improving the spatial accessibility of healthcare in North Kivu, Democratic Republic of Congo. *Applied Geography* **121**, 102262 (2020).
  11. Bondarenko Maksym, Kerr David, Sorichetta Alessandro, Tatem Andrew. Census/projection-disaggregated gridded population datasets, adjusted to match the corresponding UNPD 2020 estimates, for 51 countries across sub-Saharan Africa using building footprints. (2020).
  12. Schiavina, Marcello; Freire, Sergio; MacManus, Kytt (2019): GHS population grid multitemporal (1975, 1990, 2000, 2015) R2019A. European Commission, Joint Research Centre (JRC) DOI: 10.2905/42E8BE89-54FF-464E-BE7B-BF9E64DA5218 PID:<http://data.europa.eu/89h/0c6b9751-a71f-4062-830b-43c9f432370f>
  13. Center for International Earth Science Information Network - CIESIN - Columbia University. Gridded Population of the World, Version 4 (GPWv4): Population Count Adjusted to Match 2015 Revision of UN WPP Country Totals, Revision 11. NASA Socioeconomic Data and Applications Center (SEDAC) (2018).
  14. WorldPop (www.worldpop.org - School of Geography and Environmental Science, University of Southampton; Department of Geography and Geosciences, University of Louisville; Departement de Geographie, Universite de Namur) and Center for International Earth Science Information Network (CIESIN), Columbia University (2018). Global High Resolution Population Denominators Project - Funded by The Bill and Melinda Gates Foundation (OPP1134076)
  15. Facebook Connectivity Lab and Center for International Earth Science Information Network - CIESIN - Columbia University. High Resolution Settlement Layer (HRSL). In: *Source imagery for HRSL @ 2019 DigitalGlobe*. <https://data.humdata.org/dataset/highresolutionpopulationdensitymaps> (2019).
  16. Rose Amy N., McKee Jacob J., Sims Kelly M., Bright Edward A., Reith Andrew E., Urban Marie L. LandScan 2019. 2019 edn. Oak Ridge National Laboratory (2020). <https://landscan.ornl.gov/>
  17. Maina Joseph, *et al.* A spatial database of health facilities managed by the public health sector in sub Saharan Africa. *Scientific data* **6**, 1-8 (2019).

18. Thematic Research Network on Data and Statistics (TReNDS). Leaving no one of the map: A guide for gridded population data for sustainable development.(2020).
19. Leyk Stefan, *et al.* The spatial allocation of population: A review of large-scale gridded population data products and their fitness for use. *Earth System Science Data* **11**, (2019).
20. Yin Xu, Li Peng, Feng Zhiming, Yang Yanzhao, You Zhen, Xiao Chiwei. Which Gridded Population Data Product Is Better? Evidences from Mainland Southeast Asia (MSEA). *ISPRS International Journal of Geo-Information* **10**, 681 (2021).
21. Joseph Noel K, *et al.* Spatial access inequities and childhood immunisation uptake in Kenya. *BMC Public Health* **20**, 1-12 (2020).
22. Cairo Sarah B, *et al.* Geospatial Mapping of Pediatric Surgical Capacity in North Kivu, Democratic Republic of Congo. *World journal of surgery*, 1-9 (2020).
23. Macharia Peter M, Ouma Paul O, Gogo Ezekiel G, Snow Robert W, Noor Abdisalan M. Spatial accessibility to basic public health services in South Sudan. *Geospatial health* **12**, 510 (2017).
24. Dotse-Gborgbortsi Winfred, Tatem Andrew J, Alegana Victor, Utazi C Edson, Ruktanonchai Corrine Warren, Wright Jim. Spatial inequalities in skilled attendance at birth in Ghana: a multilevel analysis integrating health facility databases with household survey data. *Tropical Medicine & International Health*, (2020).
25. Chen Yi No, Schmitz Michelle M, Serbanescu Florina, Dynes Michelle M, Maro Godson, Kramer Michael R. Geographic access modeling of emergency obstetric and neonatal care in Kigoma Region, Tanzania: transportation schemes and programmatic implications. *Global Health: Science and Practice* **5**, 430-445 (2017).
26. Curtis Andrew, *et al.* National optimisation of accessibility to emergency obstetrical and neonatal care in Togo: a geospatial analysis. *BMJ open* **11**, e045891 (2021).
27. dos Anjos Luis António, Cabral Pedro. Geographic accessibility to primary healthcare centers in Mozambique. *International journal for equity in health* **15**, 173 (2016).
28. Hierink Fleur, Rodrigues Nelson, Muñoz Maria, Panciera Rocco, Ray Nicolas. Modelling geographical accessibility to support disaster response and rehabilitation of a healthcare system: an impact analysis of Cyclones Idai and Kenneth in Mozambique. *BMJ open* **10**, e039138 (2020).
29. Kuupiel Desmond, *et al.* Geographic accessibility to public health facilities providing tuberculosis testing services at point-of-care in the upper east region, Ghana. *BMC public health* **19**, 718 (2019).
30. Reed Fennis J, Gaughan Andrea E, Stevens Forrest R, Yetman Greg, Sorichetta Alessandro, Tatem Andrew J. Gridded population maps informed by different built settlement products. *Data* **3**, 33 (2018).
31. Smith Andrew, Bates Paul D, Wing Oliver, Sampson Christopher, Quinn Niall, Neal Jeff. New estimates of flood exposure in developing countries using high-resolution population data. *Nature communications* **10**, 1-7 (2019).
32. Bustos Maria Francisca Archila, Hall Ola, Niedomysl Thomas, Ernstson Ulf. A pixel level evaluation of five multitemporal global gridded population datasets: A case study in Sweden, 1990–2015.

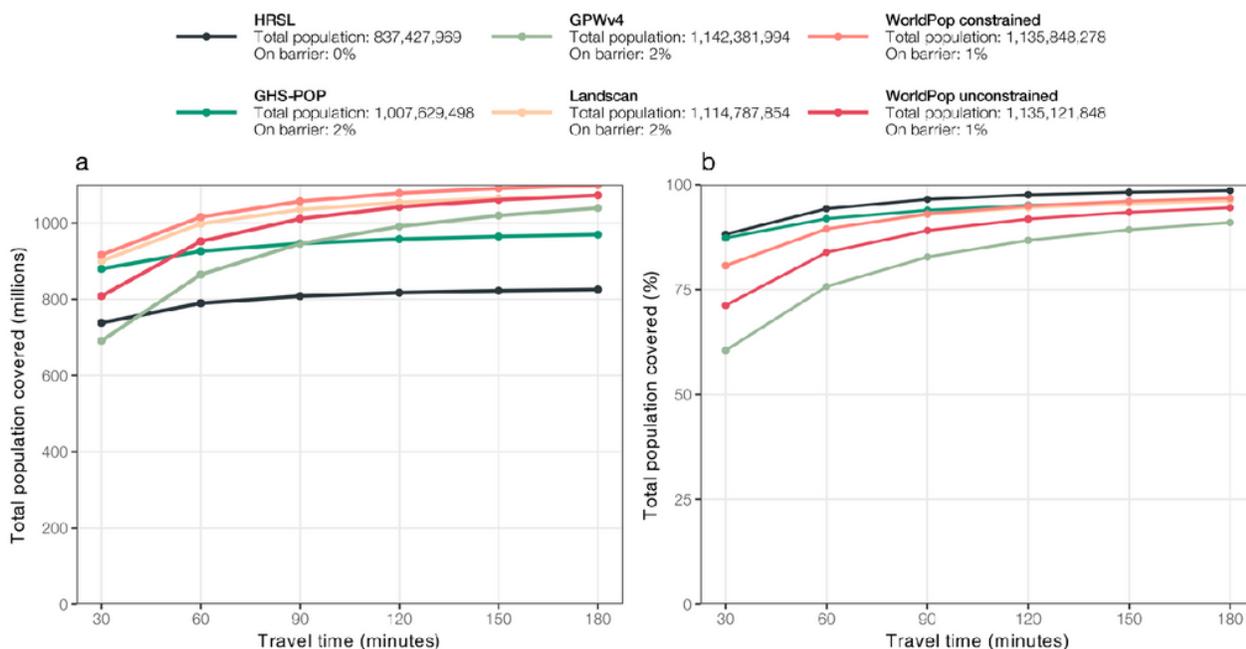
- Population and environment*, 1-23 (2020).
33. Tan Jianxia, Wang Xiuli, Pan Jay. The effect of population distribution measures on evaluating spatial accessibility of primary health-care institutions: A case study from China. *Geospatial Health* **16**, (2021).
  34. Weiss DJ, *et al.* Global maps of travel time to healthcare facilities. *Nature Medicine* **26**, 1835-1838 (2020).
  35. Palacios-Lopez Daniela, *et al.* New Perspectives for Mapping Global Population Distribution Using World Settlement Footprint Products. *Sustainability* **11**, 6056 (2019).
  36. Stevens Forrest R., Gaughan Andrea E., Linard Catherine, Tatem Andrew J. Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. *PLOS ONE* **10**, e0107042 (2015).
  37. Calka Beata, Bielecka Elzbieta. Reliability Analysis of LandScan Gridded Population Data. The Case Study of Poland. *ISPRS Int J Geo Inf* **8**, 222 (2019).
  38. Calka Beata, Bielecka Elzbieta. GHS-POP Accuracy Assessment: Poland and Portugal Case Study. *Remote Sensing* **12**, 1105 (2020).
  39. Ali Disha, *et al.* A cost-effectiveness analysis of traditional and geographic information system-supported microplanning approaches for routine immunization program management in northern Nigeria. *Vaccine* **38**, 1408-1415 (2020).
  40. Linard Catherine, Gilbert Marius, Snow Robert W, Noor Abdisalan M, Tatem Andrew J. Population distribution, settlement patterns and accessibility across Africa in 2010. *PloS one* **7**, e31743 (2012).
  41. Wardrop N. A., *et al.* Spatially disaggregated population estimates in the absence of national population and housing census data. *Proceedings of the National Academy of Sciences* **115**, 3529-3537 (2018).
  42. Nilsen Kristine, *et al.* A review of geospatial methods for population estimation and their use in constructing reproductive, maternal, newborn, child and adolescent health service indicators. *BMC health services research* **21**, 370 (2021).
  43. Alegana Victor A, *et al.* National and sub-national variation in patterns of febrile case management in sub-Saharan Africa. *Nature communications* **9**, 1-7 (2018).
  44. Messina Jane P., *et al.* The current and future global distribution and population at risk of dengue. *Nature Microbiology* **4**, 1508-1515 (2019).
  45. Utazi C. Edson, *et al.* Mapping vaccination coverage to explore the effects of delivery mechanisms and inform vaccination strategies. *Nature Communications* **10**, 1633 (2019). doi: 10.1038/s41467-019-09611-1. PMID: 30967543; PMCID: PMC6456602.
  46. Utazi CE, Tatem AJ. Precise mapping reveals gaps in global measles vaccination coverage. *Nature*. 2021 Jan;589(7842):354-355. doi: 10.1038/d41586-020-03391-1. PMID: 33328642.
  47. World Health Organization. Ending Preventable Maternal Mortality (EPMM): A renewed focus for improving maternal and newborn health and wellbeing. (2021). Available

from:<https://apps.who.int/iris/bitstream/handle/10665/350834/9789240040519-eng.pdf?sequence=1&isAllowed=y>

48. Keyes Emily B, Parker Caleb, Zissette Seth, Bailey Patricia E, Augusto Orvalho. Geographic access to emergency obstetric services: a model incorporating patient bypassing using data from Mozambique. *BMJ global health* **4**, e000772 (2019).
49. Mondal Pinki, Tatem Andrew J. Uncertainties in Measuring Populations Potentially Impacted by Sea Level Rise and Coastal Flooding. *PLoS one* **7**, e48191 (2012).
50. Brandt Cyril, De Herdt Tom. On the Political Economy of Data Collection: Lessons from the Unaccomplished Population Census (DR Congo, 2006-2018). *Secure Livelihoods Research Consortium Working Papers* **72**, (2019).
51. Boo Gianluca, *et al.* High-resolution population estimation using household survey data and building footprints. *Nature Communications* **13**, 1330 (2022).
52. Boo G, *et al.* 2021. Modelled gridded population estimates for the HautKatanga, Haut-Lomami, Ituri, Kasaï, Kasaï-Oriental, Lomami and Sud-Kivu provinces in the Democratic Republic of the Congo (2021), version 3.0. WorldPop, University of Southampton, Flowminder Foundation, École de Santé Publique de Kinshasa, Bureau Central du Recensement and Institut National de la Statistique. doi:10.5258/SOTON/WP00720
53. Lloyd Christopher T, Sorichetta Alessandro, Tatem Andrew J. High resolution global gridded data for use in population studies. *Scientific data* **4**, 1-17 (2017).
54. Dijkstra Edsger W. A note on two problems in connexion with graphs. *Numerische mathematik* **1**, 269-271 (1959).
55. Gilardi A. , Lovelace R. osmextract: Download and Import Open Street Map Data Extracts.R package version 0.3.1 edn (2021).
56. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing (2021).
57. Buchhorn M , *et al.* Copernicus Global Land Service: Land Cover 100m, epoch "2019", Globe. (2020). doi:10.5281/zenodo.3939050
58. GADM. Global Administrative Areas (GADM): maps and data. (2020).
59. OpenStreetMap contributors. Planet dump retrieved from <https://planet.osm.org>. (2021).
60. Microsoft Corporation, Weston S. doParallel: Foreach Parallel Adaptor for the 'parallel' Package. R package version 1.0.16 edn (2020).
61. Microsoft Corporation, Weston S. foreach: Provides Foreach Looping Construct. R package version 1.5.1 edn (2020).
62. Hijmans RJ. terra: Spatial Data Analysis. In: <https://CRAN.R-project.org/package=terra>. R package version 1.4-11 edn (2021).
63. OpenStreetMap Wiki. Key: highway. (2021). Available from:<https://wiki.openstreetmap.org/wiki/Key:highway>

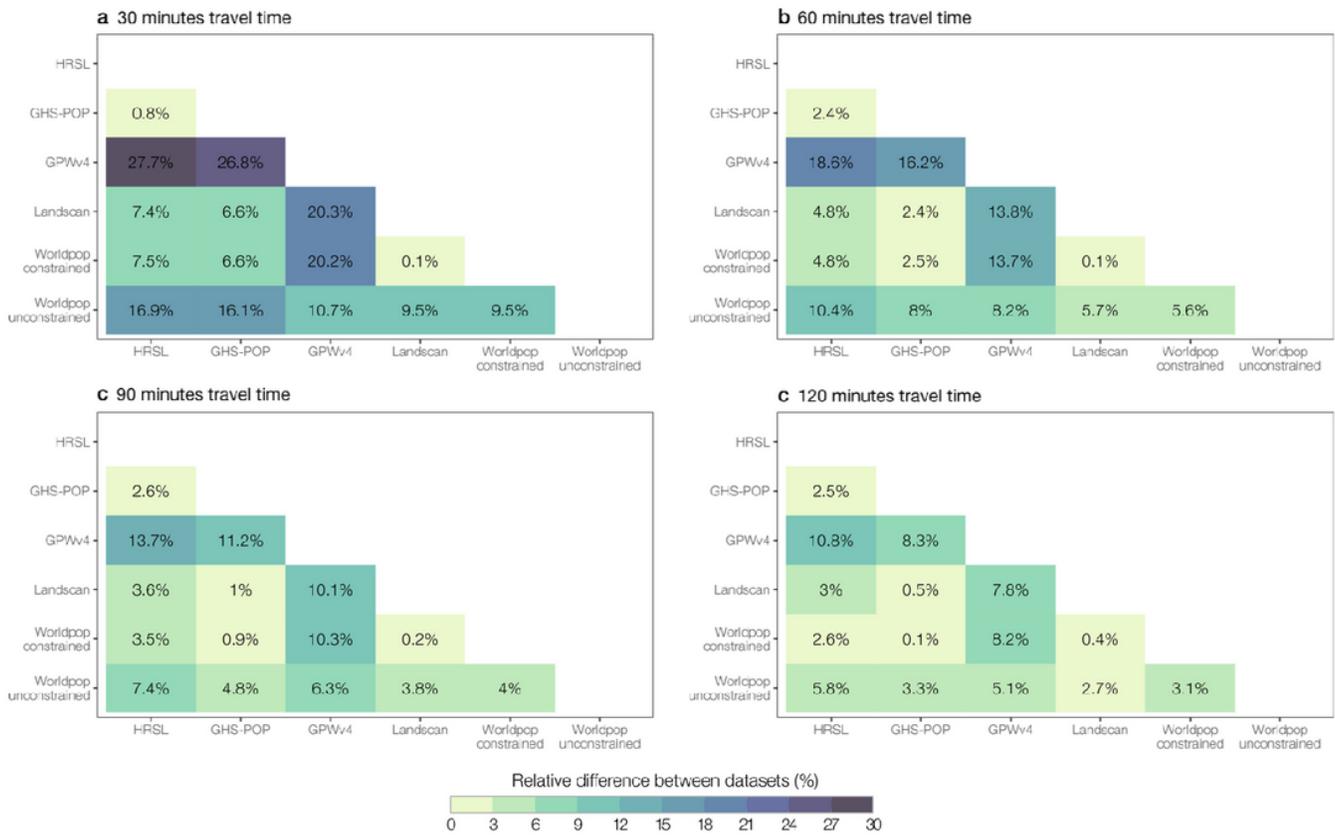
64. Hierink Fleur, Margutti Jacopo, van den Homberg Marc, Ray Nicolas. Constructing and validating a transferable epidemic risk index in data scarce environments using open data: a case study for dengue in the Philippines. *PLoS neglected tropical diseases* **16**, e0009262 (2022).
65. Ouma Paul, Macharia Peter M., Okiro Emelda, Alegana Victor. Methods of Measuring Spatial Accessibility to Health Care in Uganda. In: *Practicing Health Geography: The African Context* (ed Makanga PT). Springer International Publishing (2021).
66. Thomson Dana R, Gaughan Andrea E, Stevens Forrest R, Yetman Gregory, Elias Peter, Chen Robert. Evaluating the accuracy of gridded population estimates in slums: A case study in Nigeria and Kenya. *Urban Science* **5**, 48 (2021).

## Figures



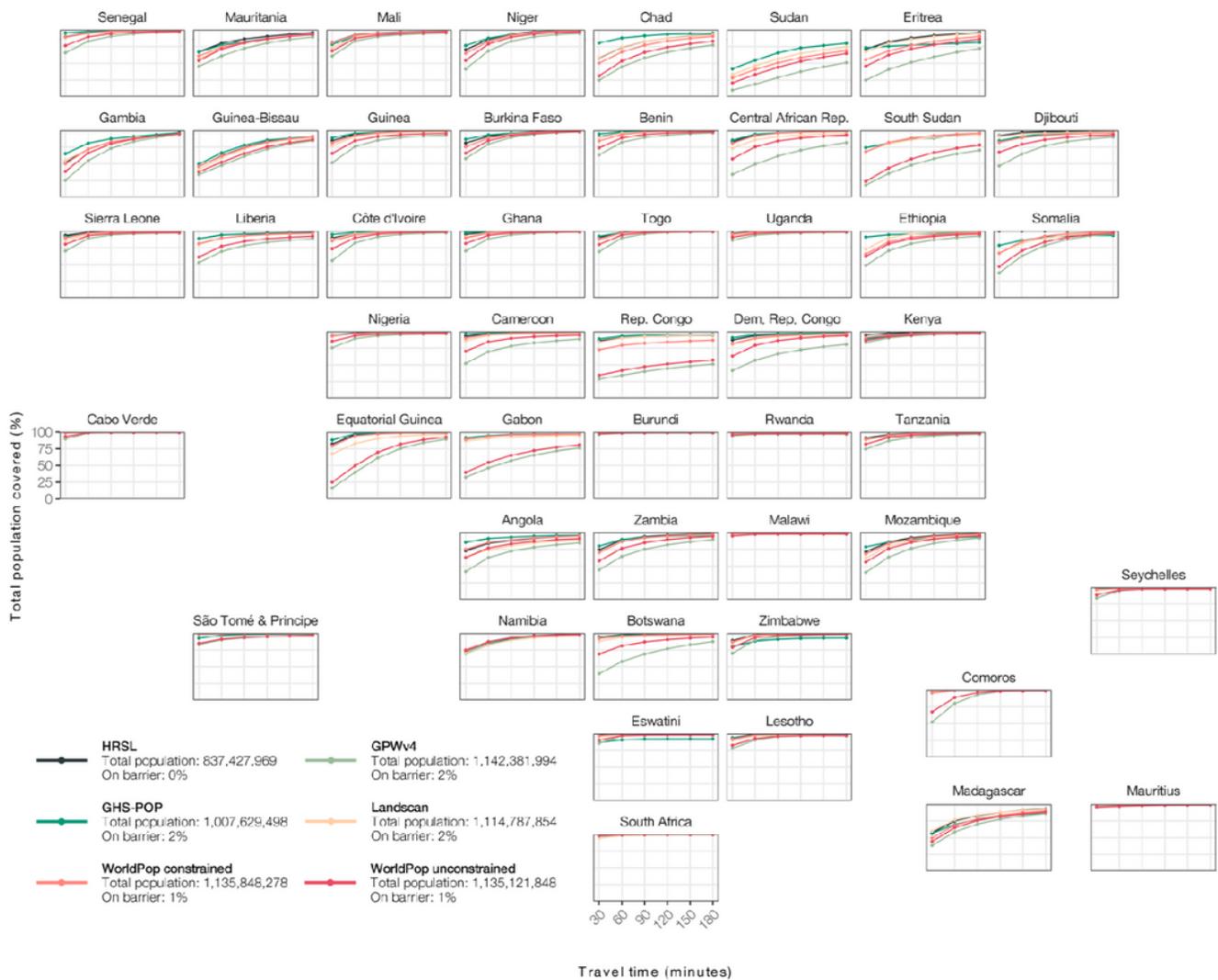
**Figure 1**

**Accessibility coverage at the sub-Saharan African level.** Absolute (a) and relative (b) continental accessibility coverage for the six different gridded population datasets: HRSL, GHS-POP, GPWv4, Landscan, WorldPop top-down constrained, WorldPop top-down unconstrained. Total population is lower for HRSL because Ethiopia, Somalia, Sudan, and South Sudan are not included in the dataset released in 2018. Legend indicates the total population falling on barriers (i.e. permanent waterbodies) and thus not included in the analysis.



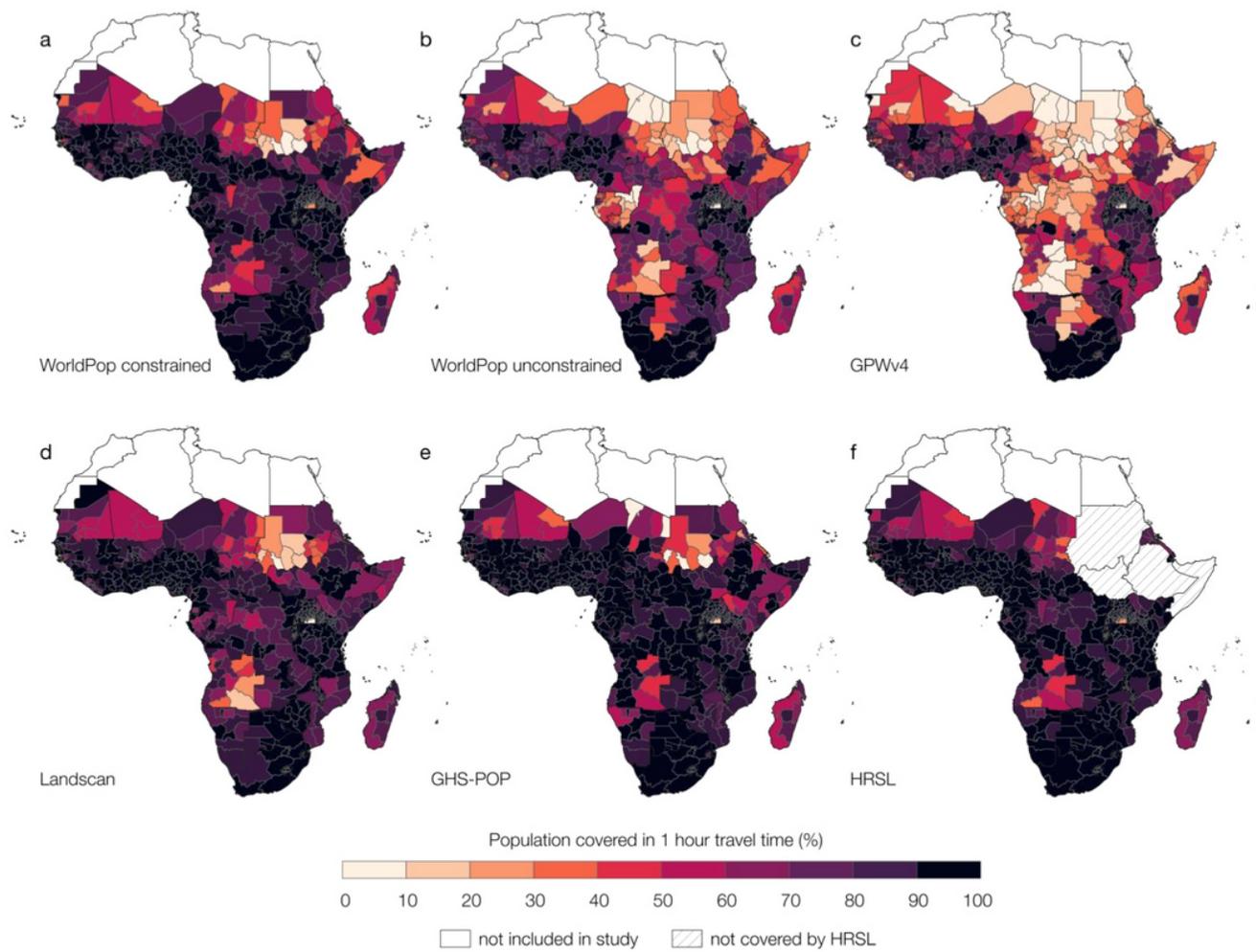
**Figure 2**

**Relative difference in accessibility coverage estimates.** The matrix shows the relative difference in accessibility coverage statistics at a) 30, b) 60, c) 90, and d) 120 minutes travel time for the six gridded population datasets for the full Sub-Saharan African region.



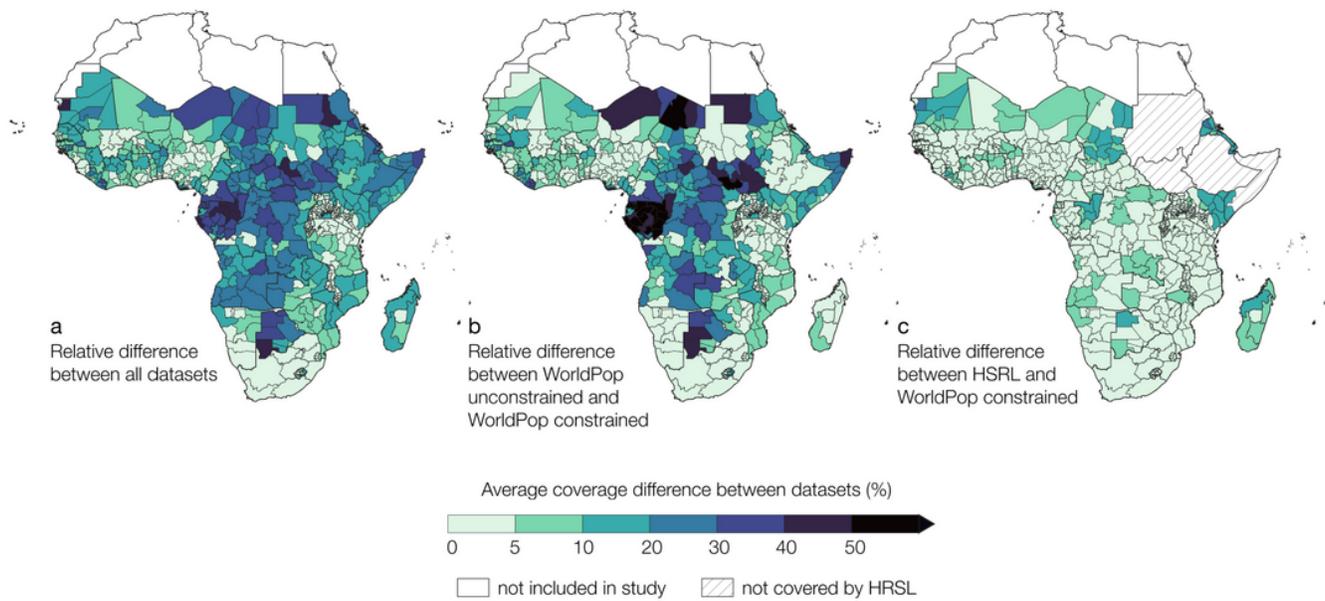
**Figure 3**

**Accessibility coverage for all sub-Saharan African countries.** National plots for all sub-Saharan African countries comparing relative accessibility coverage statistics at 30, 60, 90, 120, 150, and 180 minutes travel time for the six gridded population datasets. Each plot corresponds to the relative geographical location of the country. Legend indicates the total population falling on barriers (i.e. permanent waterbodies) and thus not included in the analysis.



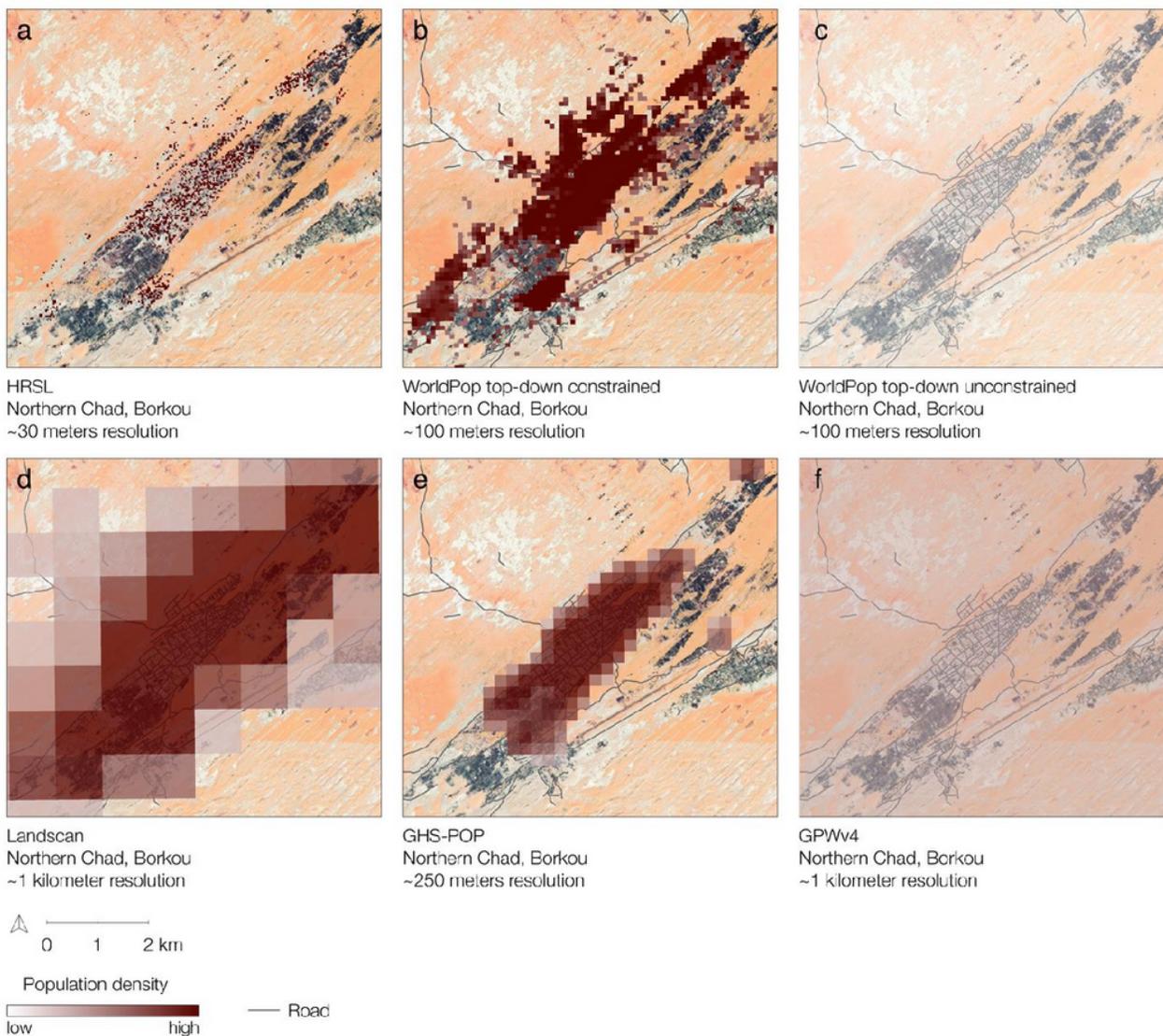
**Figure 4**

**Sub-national accessibility coverage maps for sub-Saharan Africa.** Relative accessibility coverage at a 1-hour travel time limit for a) WorldPop top-down constrained, b) WorldPop top-down unconstrained, c) GPWv4, d) Landscan, e) GHS-POP, and f) HRSL. Boundaries reflect administrative level 1.



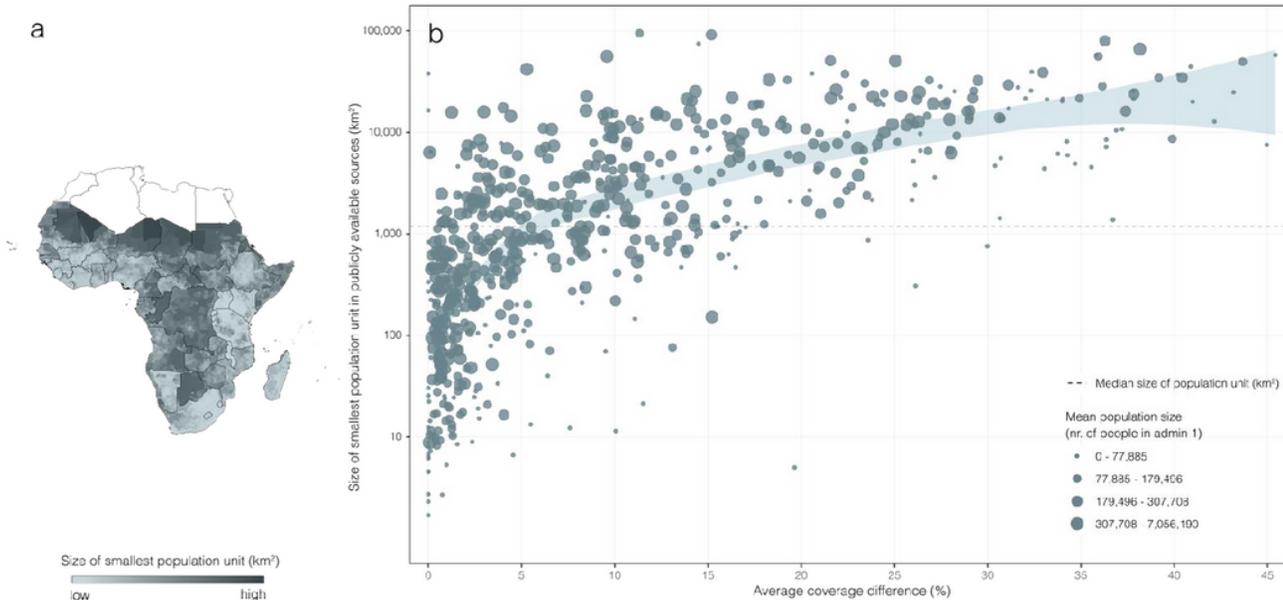
**Figure 5**

**Maps of relative difference in accessibility coverage estimates.** Maps show the average relative difference in accessibility coverage statistics at 1-hour travel time between a) all datasets, b) WorldPop top-down unconstrained and WorldPop top-down constrained, and c) HSRL and WorldPop top-down constrained for full sub-Saharan Africa.



**Figure 6**

**Visual comparison of gridded population datasets.** Visual differences between a) HRSL, b) WorldPop top-down constrained, c) WorldPop top-down unconstrained, d) Landscan, e) GHS-POP, and f) GPWv4 for Borkou, a northern region in Chad. Google satellite imagery (2022) as background. White transparent color represents low numbers of population density.



**Figure 7**

**Association between size of population unit and difference in accessibility coverage.** a) Spatial variation of the size of the smallest publicly available population unit. b) Association between population input unit size and average relative difference in accessibility coverage between the datasets. The size of the points indicates the total population averaged for all datasets at administrative level 1.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarydata2Travelscenario.xlsx](#)
- [Supplementarydata1Coveragestatisticsadmin2.xlsx](#)
- [Supplementaryfigurestable.docx](#)