Fact file

Benefits of working with probabilistic population data: Using uncertainty measures

Flowminder and WorldPop at the University of Southampton work with countries to apply hybrid census approaches and provide them with accurate high-resolution population data. To estimate populations, WorldPop has developed a novel bottom-up modelling approach, based on a Bayesian model. This method combines existing population data with satellite imagery to create **100 x 100m gridded population estimates** and **measures of uncertainty** for the whole country or targeted areas.

For more information about our Bayesian model, please refer to our fact file: “The bottom-up modelling approach”.

Our modelling produces population estimates for 100 x 100m areas called ‘grid cells’. For each grid cell, the model predicts a range of potential population densities based on the input data. The **mean value** of those densities is the best estimate of the population for that area. We also produce information on the uncertainty attached to each estimate. These **uncertainty measures** provide upper and lower figures around the mean estimate and indicate the degree of confidence in each estimate.

The **degrees of confidence**, known as confidence intervals, tell us the range of values that the population is likely to fall within at a given level of confidence (e.g. 95%). There are two types of confidence intervals that are most useful:

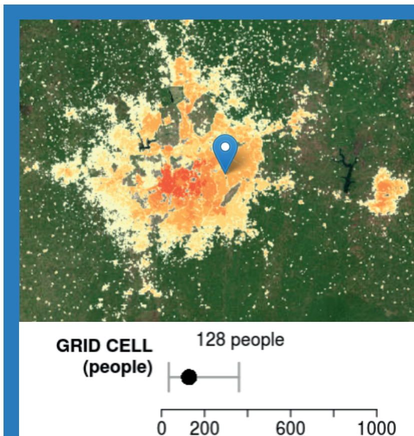
Two-tailed confidence intervals
The range of values that we can be 95% certain the population falls within

One-tailed confidence intervals
The threshold value that we can be 95% certain the population does not exceed

Why does uncertainty matter?

Observations and models cannot capture the full complexity of our world. Population estimates can never be perfectly accurate, so it can be easy to forget that they are estimates if they reported as only a single number in each grid cell.

Uncertainty in population estimates comes from limited data input, such as insufficient population data (i.e. small sample size), sampling bias that under-represents some populations (e.g. slum and informal settlement dwellers), and outdated maps of settled areas. This applies to all population estimates. Using a Bayesian method enables the consideration of these errors during the population estimation, and provides estimates with uncertainty measures.



Example of a population estimate at grid cell level

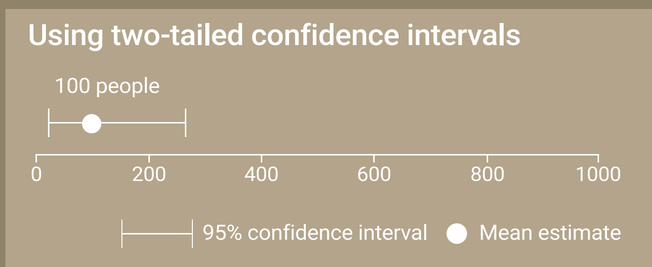
The uncertainty associated with gridded population estimates varies from place to place. For example, less uncertainty is expected in areas where microcensus surveys have been conducted than in regions with no survey data. However, even in areas where surveys are carried out, uncertainty can still be relatively high. From one neighbourhood to the other, population densities can change drastically if patterns are not captured by predictors in the statistical model (e.g. settlement type or school densities).

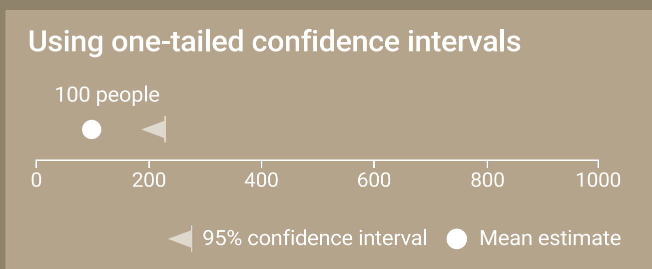
These uncertainty predictions around the mean estimates **are valuable in making more informed decisions.** For example, it is critical when determining where to build a new school, so that every child has access to quality education. Population and uncertainty estimates can also be applied to other types of operations, such as vaccination interventions, election polls locations, investment decisions about hospitals, and transportation and communication networks.

Scenario: Using uncertainty for decision-making in vaccination interventions

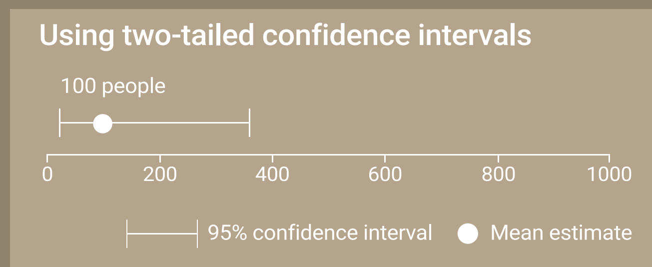
Predicting the uncertainty associated with our population estimates is critical to allocate the necessary amount of vaccines to a target population. Imagine planning a vaccination intervention in two areas using the population estimates displayed below.

Area A





Area B



If planning was solely based on the grid cell estimates above, **100** vaccines would be allocated for each area.

Incorporating the confidence intervals into the planning process, however, suggests that administering **250** vaccines for area A and **380** for area B would be a better intervention (**two-tailed confidence intervals**), to ensure enough vaccines are available to cover everyone, because it is 95% certain that the populations are within these numbers in each grid cell.

For a vaccination campaign, it would be more suitable to use the **one-tailed confidence interval**, because of the risk of having a population that exceeds the number of vaccines. In this case, the campaign would require **210** and **300** vaccines for area A and B, respectively.

Working with uncertainty measures: Proven applications

Uncertainty measures help make more informed decisions. They have proven to be particularly useful in a wide range of applications, including:

Using the upper confidence interval for vaccination planning or other logistical planning (e.g. bed nets).

Mapping model uncertainty to identify where to conduct additional microcensus surveys to improve the model.

Mapping model uncertainty to try to identify important predictors that may be missing from the model that could explain population patterns in areas with higher uncertainty.

info@flowminder.org www.flowminder.org wp@worldpop.uk www.worldpop.org