SAPIENs Guidance Document

March 2021

UKaid
from the British people

Ascend
West and Central Africa

SCI Foundation

IMPROVING HEALTH
UNLOCKING POTENTIAL

SCI Foundation

Version 2, March 2021.

Authored by Alexandra Carlin, SCI Foundation, London UK.

Supported by Ascend West and Central Innovation Fund.

IMPROVING HEALTH
UNLOCKING POTENTIAL

SCI Foundation

SAPIENs Guidance Document: A Method for Evaluating Populations When Planning for Mass Drug Administrations

Contents

1. Abbreviations and Definitions

Administrative Unit (Admin) - Territorial units which a country is divided into

Adults – Individuals above the age of 15 years old

DQA – Data Quality Assessment

DRC – Democratic Republic of Congo

ESPEN - Expanded Specials Project for the Elimination of NTDs

GIS – Geographic Information System

HDX – Humanitarian Data Exchange

IU – Implementation Unit

JAP – Joint Application Package

JRSM – Joint Request for Selected PC Medicines

JRF – Joint Reporting Forms

JSON – Java Script Open Notation

LF – Lymphatic filariasis

MDA – Mass Drug Administration

MoH – Ministry of Health

NGO – Non-Governmental Organisation

NSO – National Statistics Office

NTD – Neglected Tropical Disease(s)

OCHA – Office for the Coordination of Humanitarian Affairs

PC – Preventive Chemotherapy

Pre-SAC – Pre-school aged children (>5 years old)

SAC – School-aged Children (5-14 years old)

UNFPA – United Nations Population Fund

WHO – World Health Organisation

WOPR - WorldPop Open Population Repository

2. Introduction

The World Health Organisation (WHO) recommends specific coverage targets for preventive chemotherapy (PC) of neglected tropical diseases (NTDs) to ensure the effectiveness of mass drug administrations (MDA)[1]. Failure to reach these targets, even in a small number of infected individuals, can perpetuate transmission, undermining attempts at elimination. Therefore, accurate figures for total populations, and populations disaggregated by age and sex, at current and future implementation units (IU) are key to plan and implement PC intervention strategies.

Due to infrequent national censuses in many countries endemic for NTDs, population data available to Ministries of Health (MoH) can often be inaccurate. However, without a more reliable population source or evidence of the impact of inaccurate data, it can be difficult to address the issues.

To address this problem a project was designed named SAPIENs: A tool to conduct Small Areas Population Evaluations, funded by the Ascend West and Central Innovation Fund, which investigated modelled population data and developed a methodology to aid NTD Programme Managers and/or Data Managers in assessing population data used within an NTD Programme. This SAPIENs Guidance Document was a SAPIENs Project Output.

2.1. Purpose of the SAPIENs Guidance Document

Where expected inaccuracies or differences exist in an NTD Programmes population figures, the SAPIENs Guidance Document is available to aid NTD Programme Managers to evaluate and address concerns with official population data used for MDA planning and/or evaluation. The SAPIENs Guidance Document describes a method and relevant resources to compare and assess current population data against three other population sources, including modelled population data. Then providing further guidance on dissemination of results through a workshop, alongside discussion points to conclude on next steps for NTD Programme moving forward.

In some instances, the Democratic Republic of Congo (DRC) has been used as the example to illustrate the data collection processes. This country was picked at random from the modelled population data available at the time, and any result exemplified in the SAPIENs Guidance Document are not associated with DRC and were created using fabricated data for the purpose of this document.

[1] WHO (2011) Helminth control in school-aged children: A guide for managers of control programmes.

The SAPIENs Guidance Document includes the following process:

- **Introduction** – Outlines the modelled population data which will be used in this SAPIENs Guidance Document and their limitations
- **Phase 1: Data Preparation** – Introduces geographic information systems (GIS) data and the Population Review Workbook, and guides the collection and preparation necessary to complete the Population Review Workbook
- **Phase 2: Data Collection & Management** – Guides the download of the population data from all sources and how to input collected data into the Population Review Workbook
- **Phase 3: Data Analysis & Visualisations** – Explains how to interpret the analyses and visualisations produced by the Population Review Workbook
- **Phase 4: Data Dissemination** – Provides suggestions on how and what to share when disseminating findings to policy and/or decision makers

Many of the webpages or online resources mentioned in this document are only available in English. Where possible, guidance on how to change the language of the webpages has been provided. However, where not possible, it is recommended to use the auto-translate option in Google Chrome to convert webpages into the language of your choice. Guidance on how to auto-translate your webpage using Google Chrome can be found here or through this link bit.ly/3b7XLGP.

2.2. Target Audience

The SAPIENs Guidance Document is written for NTD Programme Managers or Data Managers working in national MoH. The instructions within can be used, and may prove useful, to other interested third parties such as, programme managers within NTD-specific non-governmental organisations (NGO), and researchers. However, please note that third parties may find it difficult to obtain some data mentioned in the SAPIEN Guidance Document if they are not available online, for example the Joint Request for Selected PC Medicines (JRSM). In these instances, some guidance may be provided on alternative sources for collection, however, as third parties are not the focus target audience this is not guaranteed.

2.3. What is modelled population data?

WorldPop[2], a research group based in the University of Southampton, helps address issues around population data accuracy, especially in areas lacking recent national census information.

[2] Website: https://www.worldpop.org/ [Accessed 29/10/2020]

They have developed, alongside country National Statistics Offices (NSO), peer-reviewed methods of modelling geo-spatial data on population distributions and demographics.

More specifically, these models estimate populations by dividing a country into squares of 100 m^2 and inputting a predicted value for the population per square. For instance, if the information given to the model suggests the square covers a block of apartment buildings in a densely populated part of a city, the population value for that square will be high. Similarly, if the square is located in the middle of a nature reserve, an area without inhabitants, the estimated population value will be low. Then total population of an area is the sum of the population value of all the squares it contains. As the use of a square grid is not tied to any particular IU boundaries, it can be adapted quickly and easily to recalculate modelled population data if IU boundaries change i.e., when re-districting occurs. This methodological document will provide instructions on how to collect two different types of modelled population datasets from WorldPop resources: 1) bottom-up datasets and 2) top-down datasets. These datasets are created through two different statistical models:

1. **Bottom-up model** uses recent microcensus survey data and geospatial data to inform a statistical model to estimate population numbers and age/sex breakdowns across unsampled locations, with measurements of uncertainty, to fill in gaps in a census where full enumeration is not possible.

2. **Top-down model** uses current national census data and geospatial datasets, usually only available as population totals per administrative units[3], to inform a statistical model which disaggregates the national census counts to 100x100m or 1x1km grid cells.

Before moving on to further steps in this method document, it is recommended that around 30 minutes is spent reading the WorldPop webpage on Mapping Populations, found here or at https://bit.ly/3nPBh0b. This web page is only available in English, so please translate using the Google Chrome translate option mentioned in the section 2.1. Purpose of the SAPIENs Guidance Document. This page provides an in-depth explanation of what both bottom-up and top-down modelling methods mean and their associated advantages and disadvantages. Additionally, a brief decision matrix provides guidance on which modelling method is most suited to what scenario.

[3] Territorial units which a country is divided into.

2.4. Considerations when using WorldPop Modelled Data for NTD Planning

(Last Updated: 12/10/2020)

Expanding on the WorldPop webpage on Mapping Populations, found here or at https://bit.ly/3nPBh0b, this section indicates the most important considerations required when deciding which WorldPop modelled population data to use specifically for NTD planning and/or reporting.

2.4.1. Modelled 1: WorldPop Bottom-up

1. The WorldPop Open Population Repository (WOPR) can be accessed through the woprVision online platform developed by WorldPop to facilitate the download of bottom-up data and is found here or at https://bit.ly/3oS1MTY. More guidance on how to use this platform and change the language is available during Phase 2: Data Collection & Management. However, as it is commonly needed to collect input information through micro-census surveys to complete the bottom-up model, not all countries are likely to have population data available through woprVision yet. In cases where bottom-up population data does not exist, there are two options:
    a. Consider using other modelled population values provided by WorldPop, for instance top-down or the rapid population mapping tools i.e., peanutButter or with this link bit.ly/3sH6CoN.
    b. Use this SAPIENs Guidance Document to investigate and collect evidence towards issues in current population data used by an NTD programme by comparing all population data available. These results can be used as justification to contact the United Nations Population Fund (UNFPA) or WorldPop to investigate further options for population data the country may have.
2. Even though the bottom-up approach provides more updated population estimates, often by using recently collected microcensus data, it should be noted that this is not a replacement for up-to-date national census data. Bottom-up data should therefore be considered as an option of more up-to-date population data through inexpensive data collection where national census data are either not available or not current (i.e., > 3 years old).

2.4.2. Modelled 2: WorldPop Top-down

1. For top-down modelled populations, there are no data available for variance and, as a result, the upper and lower confidence intervals cannot be calculated. This differs from the bottom-up modelled data, where it is available for collection.

2. Due to the modelling method, the top-down approach relies on existing population figures as an input to the model. Therefore, the quality of the modelled population outputs are dependent on the quality of the figures being inputted. In the case of WorldPop top-down approach, this is typically national census with projections or official estimates; therefore, if these are out-dated, the modelled population outputs may still be inaccurate. However, in instances where current population data are not available at a fine geographic scale, top-down modelled estimates could provide a better insight into the current spatial distributions of populations.
3. The key strength of the top-down model is that it is available for all countries. Additionally, it is possible to change the source of population used for these models with R Programming code and guidance made available by WorldPop. However, it will require moderate to extensive experience with R coding and statistical modelling to implement.
4. In the absence of a facilitated download platform such as woprVision, typically, top-down modelled population data will be downloadable in what is called a “raster” file, which appears similar to a photo. Raster data is a matrix of cells organised into a grid, where each grid cell value represents real-world data, such as temperature. More information on raster data can be found here, or at bit.ly/2OeqFvM. To read raster files, the use of a programme (i.e., R Programming or QGIS) is required to remove the numeric values from the raster file. This process is explained in 4.4. Modelled 2: Obtaining WorldPop Top-Down data from the WorldPop website.

2.5. Extended options for Download of Modelled Data for NTD Planning

There are various methods to download WorldPop modelled data introduced in this document. To aid the understanding of which option is most appropriate to you when downloading the Modelled 1: Bottom-up and Modelled 2: Top-down a skill ladder is available in Table 1 below. This table is organised by your, or your team’s, capacity with R programming with options for “no to limited”, “moderate”, and “extensive previous experience” with R programming. This SAPIENs Guidance Document will focus on guiding an individual through the “No to limited previous experience” option as this is the most accessible. During Phase 2: Data Collection & Management, if you have the capability of proceeding with the “moderate” or “extensive previous experience”, please proceed with the data collection of Modelled 1: Bottom-Up and Modelled 2: Top-Down using the guidance and codes linked. However, please complete the rest of the SAPIENs Guidance Document as written, without further changes.

Table 1. Skill ladder representing the options available to access WorldPop Modelled Data based on R Programming capacity.

	No to limited experience with R programming	Moderate experience with R programming	Extensive experience with R programming
Modelled 1: Bottom-Up	Use the SAPIENs Guidance Method to collect bottom-up modelled populations using the woprVision interface.	Use the SAPIENs Guidance Method to collect bottom-up modelled populations using the woprVision interface.	Investigate possibility of microcensus data collection, once data is collected, modify bottom-up data model code available through the WorldPop journal articles with collected microcensus data.
Modelled 2: Top-Down	Use the SAPIENs Guidance Method to collect top-down modelled populations through the QGIS wpgp Datasets.	Use the code available through the WorldPop website to download the top-down TIFF files through R Programming. GitHub R Programming available at https://bit.ly/3tfH03p or found here.	Use a modification of the top-down code to run the model with the Population 1: Drug Requests as the preliminary data input, not the outdated national census. WorldPop code available at worldpop.org/wprfpms.

3. Phase 1: Data Preparation

For using either type of WorldPop modelled data, it is necessary to first obtain a digital vector file depicting the borders of the IUs that can be used in GIS programmes. Vector data is how geographic or spatial data is stored, comprising of lines or arcs, with start and end points which meet at nodes, creating for instance a map of administrative borders. More information on vector data can be found here or at bit.ly/3kzEOjo.

3.1. Commonly available formats of GIS data

GIS files exist in a variety of formats. A general introduction is available here on Wikipedia (bit.ly/3bRaMDD). To delineate IU borders requires a two-dimensional polygon covering the particular area of the country or countries we are interested in. This information is commonly encoded in three kinds of data files:

1. **Shapefiles:** Shapefile is the most common filetype, with the file extension ".shp".

2. **Geodatabases:** A geodatabase is a database optimized for storing data defined in a geometric space. It is commonly available for download as a zip file, containing a folder *folder_name.gdb* which encompasses the database.

3. **GeoJSON files:** GeoJSON is a file format designed for representing simple geographical features, based on the JavaScript Object Notation (JSON) format (see here for more details on the JavaScript Open Notation – JSON – format or through bit.ly/30aDgmh).

3.2. Finding GIS data

The main sources to collect GIS data from are: NSO, MoH, other national institutions, the Humanitarian Data Exchange (HDX) platform (found here or through https://data.humdata.org/fr/), or at the GADM platform (found here or through https://gadm.org/[4]).

Which source to use will depend on the context. In general, it is advised to try to you try obtaining the data first from your NTD department within the MoH. This is particularly important if the IUs used the NTD programme are sometimes internal to the NTD department you are a part of. For example, if the IU is a *health district*, this kind of district may not be used by other government ministries or agencies.

If the data are not available from within the MoH, it is advisable to try to obtain the data from other government sources in the country of interest, as they are tasked with defining the borders and updating them. For example, a countries NSO or land use administration. Certain agencies may charge fees for the access to the data.

The third option, if no official data are found, is to search for available data online, otherwise known as “open source” data. A useful aggregator of such data is the HDX platform, a data sharing platform managed by the United Nations Office for the Coordination of Humanitarian Affairs (OCHA). To access the HDX platform click here or go to https://data.humdata.org/fr/. Since this is an open platform, it is advised to check the underlying data, or “metadata”, carefully to determine how up-to-date the data may be and how reliable both the source and the data itself are.

3.3. Manipulating GIS data

3.3.1. Why manipulate GIS data

The GIS data you currently have may contain more than your selected IUs, meaning a subset of the data (e.g., only implementation units from a certain part of the country, not the whole) needs to be extracted. Alternatively, your data may be dispersed over several data files, meaning it needs to be merged into a single file.

[4] Website: https://en.wikipedia.org/wiki/GADM (English only) [Accessed on: 02/03/2021]

In addition, the data may not be in the desired format. For the collection of population sources, Modelled 1: Bottom-up and Modelled 2: Top-down, it is necessary to have the GIS data in the GeoJSON format. Therefore, if the acquired GIS data is not in the GeoJSON format you will need to convert it as indicated in section 3.3.4. How to export data in a GeoJSON format.

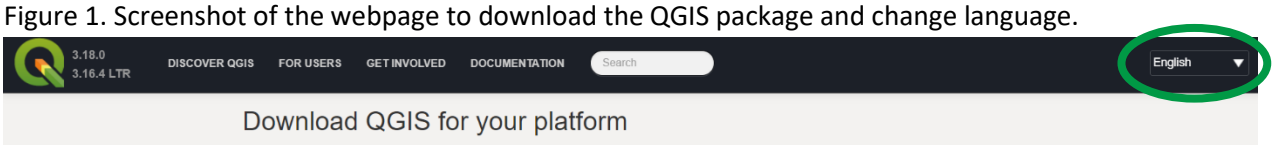
3.3.2. What free software can be used

A number of free applications are available for manipulating and using GIS data, including GIS interface programs or more generic programming languages like R or Python. For a novel user, a GIS software application is the most accessible. A comparison of programs is available here or through https://bit.ly/2LtUNm3. This website is written in English, therefore if another language is required, it is suggested to use the Google Chrome translate option mentioned in section 2.1. Purpose of the SAPIENs Guidance Document.

This document explains how to proceed using the opensource GIS software: QGIS. QGIS is a freely available GIS software package with a wide user community and good support available online. Click here, or go to here https://bit.ly/35Okdl5, to download QGIS. To change the language of the QGIS download page, please click, and change to the language of your choice using the drop-down box in the top right hand corner (green circle, Figure 1). For an introduction to QGIS a range of training videos across different languages are available here or by bit.ly/304e54V.

The official QGIS language is English; therefore, the package itself is only available in English. To add the context required for use of QGIS, French translations will be provided in brackets alongside the English words/terms.

Figure 1. Screenshot of the webpage to download the QGIS package and change language.



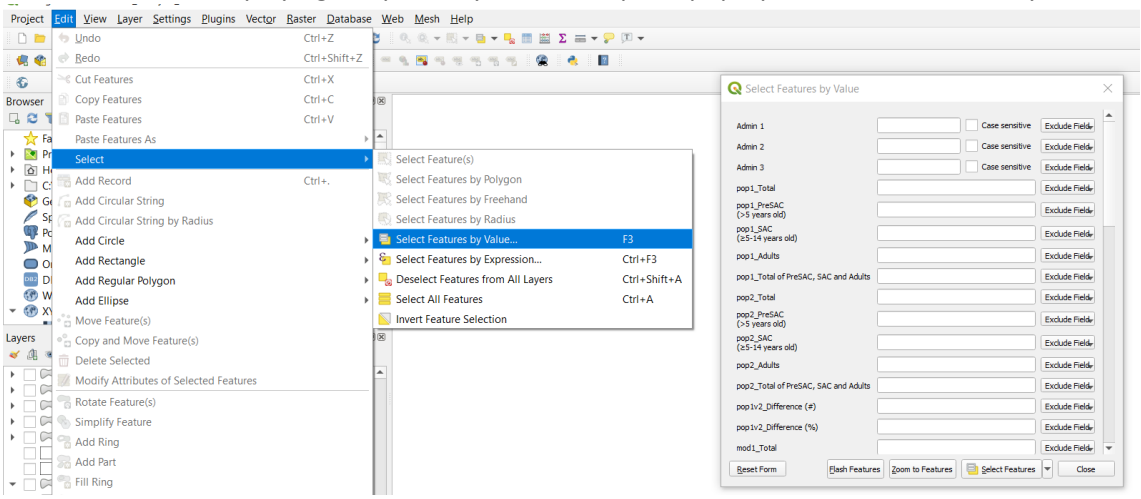
3.3.3. How to select areas in GIS data by attribute

Most commonly, a dataset for IUs will contain the IUs for the whole country. You may be interested only in certain areas. For example, if MDA is only happening in the east of the country or if the modelled data is only available for particular areas of the country, you need to reduce the data to the areas of interest.

The easiest way to select areas is using the attributes contained in the data. Suppose you have a dataset, in shapefile (.shp) format, containing the borders of all the IU in the DRC, but you are only interested in estimating population in Kongo-Central, where you are planning an upcoming MDA. Suppose further, your data has a column that identifies to which province the different IUs belong. This is commonly present in GIS files to define administrative boundaries. You can then

use the “Select Features by Value” or “Select Features by Expression” tools of QGIS to select all features, i.e., all IUs, for which the province information is your desired region.

Figure 2. A Screenshot displaying the pathway and subsequent pop-up for "Select Feature by Value".



Click here, or at bit.ly/3r7K6VC, for a QGIS tutorials from explaining how to perform this type of selection by attribute.

3.3.4. How to export data in GeoJSON format

Most administrative boundaries (GIS) data are stored as a shapefile (.shp). For further guidance, a tutorial on how to export a datafile as a GeoJSON is available here, or by bit.ly/3r7K6VC.

3.4. Preparation of SAPIENs Population Review Workbook

The population data collected is inputted into the Population Review Workbook. This Excel Workbook has been created for use alongside this SAPIENs Guidance Document to standardise data collection, ease data entry, and automate data analysis and visualisations necessary for presenting results.

A brief explanation on the sheets within the Population Review Workbook are available in Table 2 below, or on “Sheet 1. Instructions” within the Workbook.

Population_Review
_Workbook_Templa

Table 2. Table found in Population Review Workbook, Sheet: 1. Instructions, with Sheet contents and complementary Phases in SAPIENs Guidance Document.

Sheet Number	Sheet Name	Corresponding Phase in SAPIENs Guidance Document	What does it contain?
1	Instructions	NA	Includes instructions on the Population Review Workbook and collects preliminary definition tables necessary for the Data Collection to be filled by individual.
2	Data Collection	Phase 2: Data Collection & Management	Includes an empty template to collect all the raw population data collected from all population sources using the SAPIENs Guidance Document, alongside some preliminary automated calculations for analyses.
3	Data Analysis	Phase 3: Data Analysis & Visualisations, 5.1. Sheet: 2. Data Collection, 5.2. Sheet: 3. Data Analysis	Includes four results tables with automated calculations and an additional table with guidance for a manual analysis.
4	Data Visualisations	Phase 3: Data Analysis & Visualisations, 5.3. Sheet: 4. Data Visualisations	Includes for data visualisations which are automated from the data in the 3. Data Analysis sheet.
5	Admin Name Variations	Phase 2: Data Collection & Management	Contains an empty template to list of the Administrative unit names and records any variations across the documents in the names in comparison to Population 1: Drug Request.

For preparation of Phase 2: Data Collection two tables are required to be filled on Sheet: 1. Instructions of the Population Review Workbook.

- **Population Sources Record Table:** Record the official sources of all your chosen population values in the table named “Population Sources Record”. The naming of source should include the *name of the document* (i.e., JRSM or MDA Data Report), the *year of the data*, and the *platform of collection* (i.e., Expanded Specials Project for the Elimination of NTDs (ESPEN) Portal or HDX). For instance, if the Population 1: Drug Request was collected from the 2019 JRSM found on the ESPEN Portal, the source could be recorded as “JRSM Data 2019 (ESPEN Portal)”.

Figure 3. Screenshot of the Population Source Record table found in Sheet: 1. Instructions of the Population Review Workbook.

Population Sources Record		
Standardised Source Name	Sources	Website Links (if available)
Population 1: Drug Request		
Population 2: Treatment Reporting		
Modelled Population 1		
Modelled Population 2		
Number of Treatments		
Population 3: Third Party (*optional*)		

- **Programmatic Admin Unit Definitions:** Record the names of your country's administrative units in the table "Programmatic Admin Unit Definitions" included in Sheet 1: Instructions. For instance, typically this could be Admin 1: Region, Admin 2: District, Admin 3: Sub-District[5].
    - To note, if you require to go to a smaller administrative unit than Admin 3, please amend the administrative column headers as necessary (cells B14, C14 and D14), additionally manually inputting the administrative unit names (cells B15, C15, and D15).

Figure 4. Example of a Programmatic Admin Unit Definitions table (lower), and how this is displayed automatically alongside the Admin 1, Admin 2, and Admin 3 columns (upper).

Admin 1	Admin 2	Admin 3
Region	District	Sub-district
Region 1	District 1	Sub-district 1
Region 1	District 1	Sub-district 2
Region 1	District 2	Sub-district 1
Region 1	District 3	Sub-district 1

Programmatic Admin Unit Definitions		
Admin Level	Name of Admin Level	Number of Admin Units
Admin 1	Region	
Admin 2	District	
Admin 3	Sub-district	
Admin 4	*i.e. Village or Community*	
Admin 5		

4. Phase 2: Data Collection & Management

During data collection you will collect four population sources and a dataset of treatment numbers. You then have the option to collect one additional population source if evidence is required for national-level decisions, not NTD programmatic-level decisions. Guidance to collect this source is available in Annex 1: Optional Population Sources for National-level Decision Making.

1. Population 1: Drug Requests – The population source used during MDA drug request submissions.
2. Population 2: Treatment Reporting – The population source used during MDA treatment reporting, either internal (MoH) or external (ESPEN). Preference is shown toward that the population source which are used for internal MoH reporting if this differs to that of provided to ESPEN.
3. Modelled Population 1 – The results from WorldPop bottom-up modelling.

[5] Admin 1, Admin 2, and Admin 3 are typical nomenclature used for administrative hierarchical order. For instance, region, district, sub-district.

4. Modelled Population 2 – The results from WorldPop top-down modelling.
5. Population 3: Third-Party – The population source used by an alternative humanitarian aid provider, or alternative trusted body. To be used in instances of National-level Decision Making.
6. Number of Treatments – The treatment numbers from the corresponding MDA planned using Population 1: Drug Requests and Population 2: Treatment Reporting.

This collected population data will need to be:

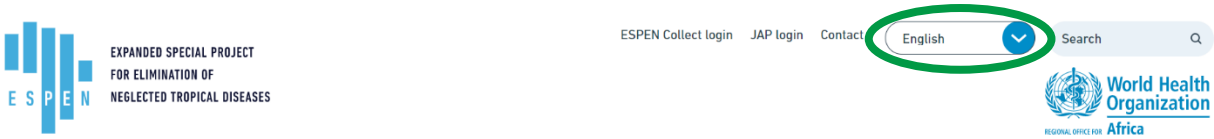
1. Population 1: Drug Requests, Population 2: Treatment Reporting, and Number of Treatments will be from the same and most recent MDA where all this data are available, and the Modelled Population 1 & 2 would be collected for the same year as the MDA. Where the MDA is conducted over two years, this would be the year of MDA start.
2. To the administrative level which coincides with your collected Shapefile found in Phase 1: Data Preparation.
    a. Note: It is common that administrative units will not 100% match when comparing different files. Therefore, find the sources with the most matching administrative units, and record the differences in spelling, name or missing values in sheet named “Admin_Name_Variations” in your Population Review Workbook. For instance, in one document a district could be name “Dingela” but another source names this “Dingila”.
3. Disaggregated by age, particularly to the age-groups required for NTD programme planning. For instance, in the case of schistosomiasis the disaggregation would be PreSAC (<5 years), SAC (5-14 years), and Adults (15 years and above).

4.1. Population 1: Drug Request

The three suggested sources for you to collect Population 1: Drug Request for NTD programmes are:

1. **ESPEN Portal**. Data from the most recent Joint Application Packages (JAP) for country NTD programme planning can often be downloaded here through the ESPEN portal (https://bit.ly/2LCKr3a). To change the language of the ESPEN portal the drop down box in the top right hand corner can be used to change the language to French, English, Portuguese or Spanish (green circle – Figure 5). The specific document which you would need is the JRSM. The JRSM is the Excel sheet which you use as a MoH NTD Programme to request medicines from WHO, therefore this is suggested as a reliable source.

Figure 5. A screenshot of the ESPEN webpage header indicating how to change the webpage language.



2. **National NTD Programmes PC Treatment Plans**. Data used to project treatment numbers of the NTD programme across a number of years.
3. **Other National NTD Programme recommended data**. Other data which may be used by your MoH for NTD Programme planning.

Please select one of these options and collect this information. It is suggested that you save this file using the name Population 1_PopSource_CCC_Year, therefore for a 2019 DRC JRSM this would be Population 1_JRSM_COD_2019. On this, "CCC" is defined as the three letter ISO Code for that country, these country specific codes are available here, or by bit.ly/3kAKQQA.

4.2. Population 2: Treatment Reporting

These are the population figures to the IU you have chosen, which are used in your internal country treatment reporting mechanisms after an MDA. This information has been typically seen collected disaggregated by age groups (Pre-SAC, SAC and Adults), and can often be referred to as the "target populations" for these MDAs. Additionally, these population data then most often serve as the basis for calculating the reported coverage for that MDA.

This target population information would be your Joint Reporting Forms (JRF) which you either have a copy of or can be accessed through the ESPEN Portal. If you are someone out with the MoH you need to request the JRF from the MoH NTD Programme of your selected country or through the Joint Reporting Forms (JRF) available through ESPEN Portal.

It is suggested that you save this file using the name Population 2_PopSource_CCC_Year. Therefore, for a 2019 DRC MDA this would be Population 2_MoH_COD_2019.

4.3. Modelled 1: Obtaining WorldPop Bottom-Up data from the woprVision web interface

This section lays out the steps to obtain the bottom-up population estimates using the woprVision platform. This will use the GeoJSON file collected during Phase 1: Data Preparation. The woprVision platform can be found here online or by using https://apps.worldpop.org/woprVision/. More details on the options the woprVision platform offers are accessible through the Help tab on the woprVision website (red circle – Figure 6).

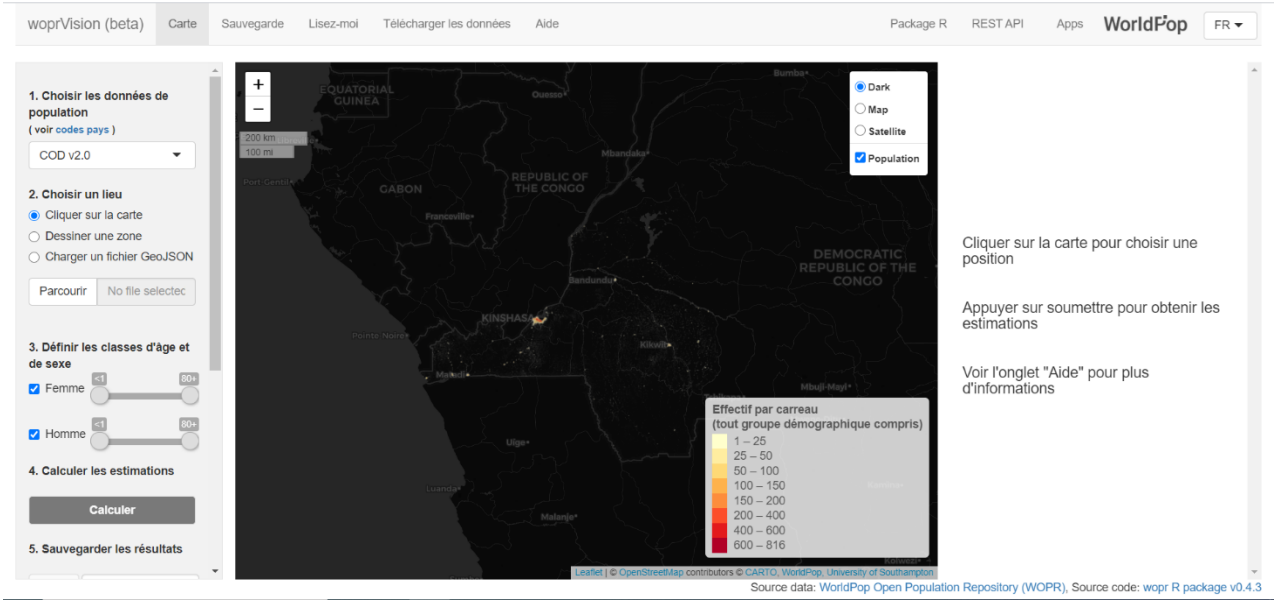
The language of woprVision can be changed by using the drop down box in the top right hand corner (green circle – Figure 6). The available languages are French, Portuguese, and Spanish.

Figure 6. Screenshot of the woprVision (beta) webpage header, indicating how to change the webpage language.

Step 1 – Overview of the site

Before commencing, it is best to explain what we see when we go on the woprVision website. Figure 7, below, shows what is available to us. Across the top there are the three tabs. By default, the website will show the *Map* tab. Additionally we may click on the *Saved* or the *Help* tab. The *Help* tab contains further explanations on the site and how to use it. The *Saved* tab compiles data investigated during a session. The goal of this guide is to export the data and thus will not concern itself with saving the data on the web and the *Saved* tab.

Figure 7 - woprVision website overview.

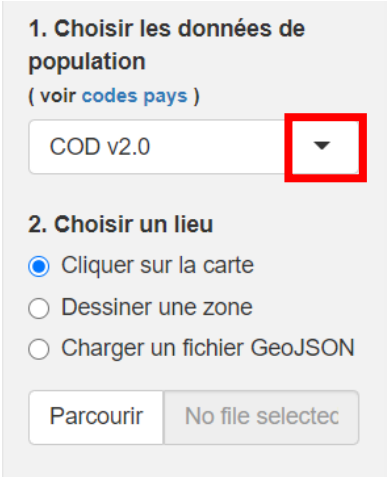


The central part shows the area of the country for which we want to obtain the data while the panel on the left-hand side contains the options to control the data settings. How to set these is showcased in the following steps.

4.3.1. Step 2 – Selecting the Country

We first need to select the country we are working on. Clicking on the downward arrow highlighted in red in Figure 8 allows to change the country. For this example, we will use the DRC, represented by its three letter ISO country code COD (https://bit.ly/3bNQEUy).[6]

Figure 8 - Selecting the country – woprVision website



4.3.2. Step 3 – Selecting the Location

The woprVision interface allows for three different forms to obtain estimates: clicking on a point in the map, drawing an area on the map known as a “polygon”, or uploading a GeoJSON file. The first two options allow us to interact directly with the map in the central panel. These are not further discussed here but are covered in the *Help* tab of the website.

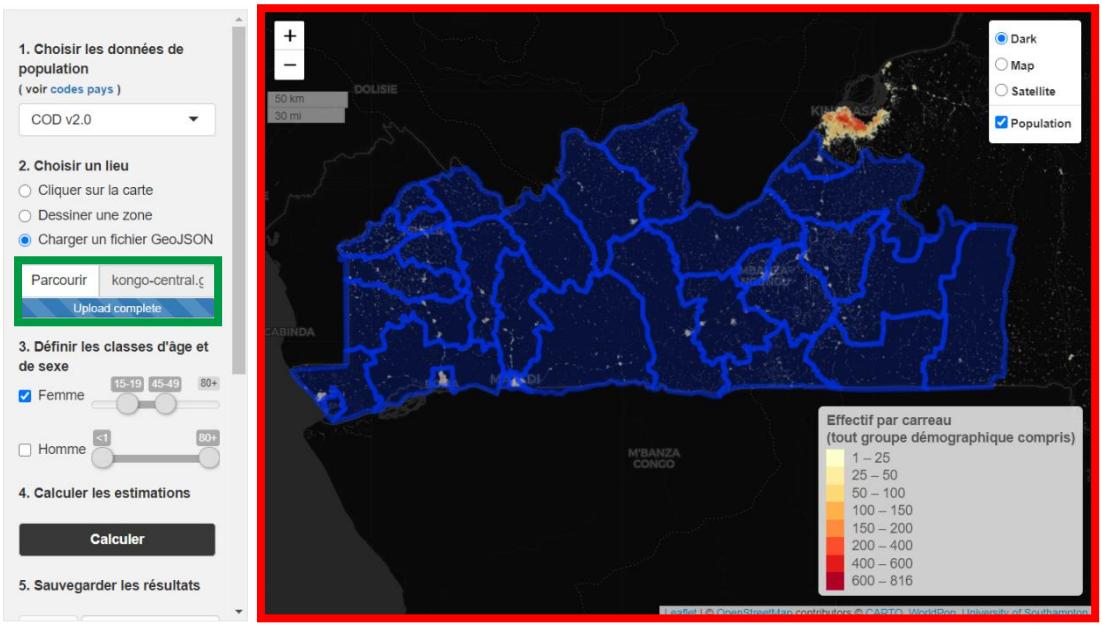
At this stage we have already obtained a digital rendering of IU boundaries (shapefile), and have converted these to GeoJSON format, using Phase 1: Data Preparation. Therefore, we select the “Upload GeoJSON file” option and browse to which file location it is stored in. To note, if you have not got this file format, please go back to complete Phase 1: Data Preparation to collect this GeoJSON file.

Once the file has been selected a bar will appear under the “Browse” box indicating the progress in updating the file to woprVision (see green box on Figure 9). Once the upload is complete the progress box will be filled, and the map will show the outlines of the administrative borders we previously selected (see red box on Figure 9). If the map does not reflect GeoJSON file content,

[6] At the time of writing (March 2021) the woprVision website provided users data for Burkina Faso, Mozambique, Sierra Leone, Ghana, Nigeria, Zambia, and five provinces of the DRC.

even if the upload has been completed, it is likely an error occurred. Check whether the file is of the correct format.[7]

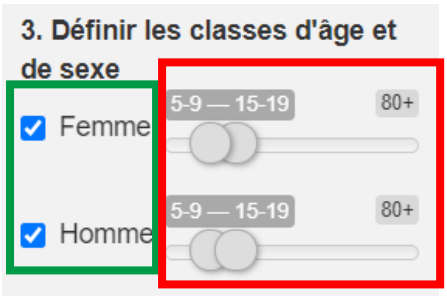
Figure 9. Selecting the location via GeoJSON file.



4.3.3. Step 4 – Selecting the age-sex groups

The next item on the settings panel is the age-sex groups for which we want to collect population estimates. We can tick and untick the sex we want to include. In Figure 10 (green box) both options have been ticked, so the population estimate will reflect the sum of both sexes. If we want information for only one sex, we need to untick the one we do not wish to see reflected in the estimates.

Figure 10 - Age-sex grouping options.

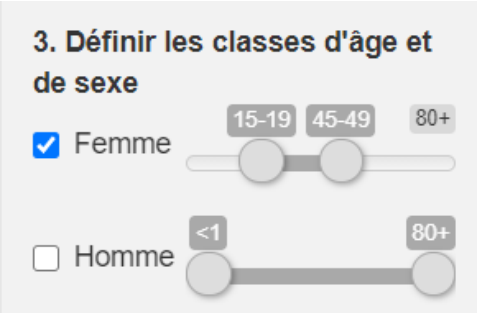


[7] One error noted during testing was that, if the GeoJSON file contained a column labelled “ID” with several “0”s (the number zero) as entries, the upload, despite complete would flag up an error and the map would not show as it does in 9. Editing the file to replace the “0”s with other values or changing the column name solved the issue.
If your file is of the correct form and format and you still experience an error, we advise contacting the WorldPop team as the woprVision site is still a beta at the time of writing and certain specifications are still being worked out.

The other part of this step is to choose the age group we are interested in. For each sex there is a sliding bar with buttons on each end to determine the ages. The buttons move along in 5-year age steps. In Figure 10 (red box) the buttons have been put so that we obtain estimates for the population aged between 5 and 19 years for both sexes. If we wanted the results for a single age group, both buttons would overlap.

As a further example, Figure 11 is set to obtain estimates for what is commonly referred to as women of child-bearing age, between 14 and 49 years of age, so excluding both younger and older women and men altogether.

Figure 11 - Selecting estimates for women of child-bearing age.



4.3.4. Step 5 – Selecting the uncertainty parameters

The woprVision website allows us not only to obtain a point estimate of the population in the area of interest for the age-sex group we wish but also to gauge the uncertainty of that estimate. Under *Options* we can first set the confidence level (red box in Figure 12). This represents the confidence we can have in the limits of the confidence interval. It is set by default at 95% and goes up to 99%. The higher the confidence level, the further the limits will be from the point estimate.[8]

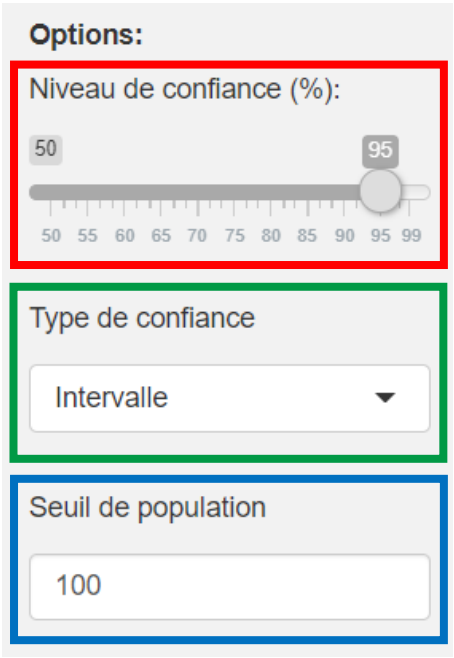
The second parameter to set is the confidence type (green box in Figure 12). There are three options available: Interval, Lower Limit, and Upper Limit. Any of these options is useable. We recommend setting it to the default “Interval” setting.[9]

The final parameter is to set a specific population threshold (blue box in Figure 12). The woprVision algorithm will calculate the probability that the population is above the inputted

[8] An introductory tutorial on the concept of confidence intervals is available on here from Khan Academy (English only).
[9] The lower and upper limit settings will still show a lower and upper limit in the results. Each can be read as the one-sided confidence interval. For example, the lower limit of the “Interval” setting at 90% confidence is equal to the lower limit of the “Lower Limit” setting at 95% confidence.

population value. Only a single threshold can be set against which the probability for all administrative units of the underlying GeoJSON will measured.

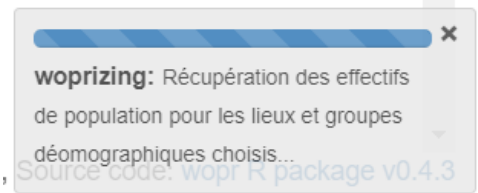
Figure 12 - Setting uncertainty parameters.



4.3.5. Step 6 – Saving the woprVision output

Once all settings have been set, we may proceed and click on the submit button on the left-hand side. After clicking it, the website will start working and the window shown in Figure 13 will appear on the lower right corner of the screen.

Figure 13 - Working pop-up after submitting data.



After the website has finished calculating the population estimates the options window shown in Figure 14 will appear. It gives us the option to save the data as a GeoJSON (red box) or as a comma separated value (CSV) spreadsheet. Click “Format CSV”.

Figure 14 – Saving output options.

The downloaded spreadsheet will contain the columns included in the GeoJSON and add columns with the woprVision estimates. The four relevant columns are *pop_mean* containing the estimated mean population value for each IU, *pop_median* containing the median of the estimated population in each IU, as well as *pop_lower* and *pop_upper*, which contain the lower and upper limits of the confidence interval at the desired confidence level (which you set in Step 4). Figure 15 shows the four columns highlighted in an example output for the DRC.

Figure 15 – Example of the four population columns of the woprVision CSV output.

	T	U	V	W	X	Y
	ADM2	pop_mean	pop_median	pop_lower	pop_upper	above
L' 3	CD2003	172203.2773	171860	150636.6	195450.45	
3' 8	CD2003	11155.357	11162.5	9614.975	12662.075	
3' 4	CD2002	196937.938	196410	172609.7	223464.3	
4' 2	CD2003	44351.9809	44582	35061.425	53169.25	
3' 0	CD2009	106006.8075	105771.5	92573.85	120668.825	
7' 1	CD2009	70755.1873	70684	61491.95	80723.25	
3' 3	CD2005	141453.9303	141292	126013.6	158005.175	
0' 1	CD2007	20376.3067	20202	17018.875	24615.125	
6,	CD2007	72051.3873	71582	62575.65	83858.025	
5' 2	CD2007	32628.6194	32498	28798.975	37198.05	
10	CD2010	38387.5034	38578.5	30205.65	45909	

4.4. Modelled 2: Obtaining WorldPop Top-Down data from the WorldPop website

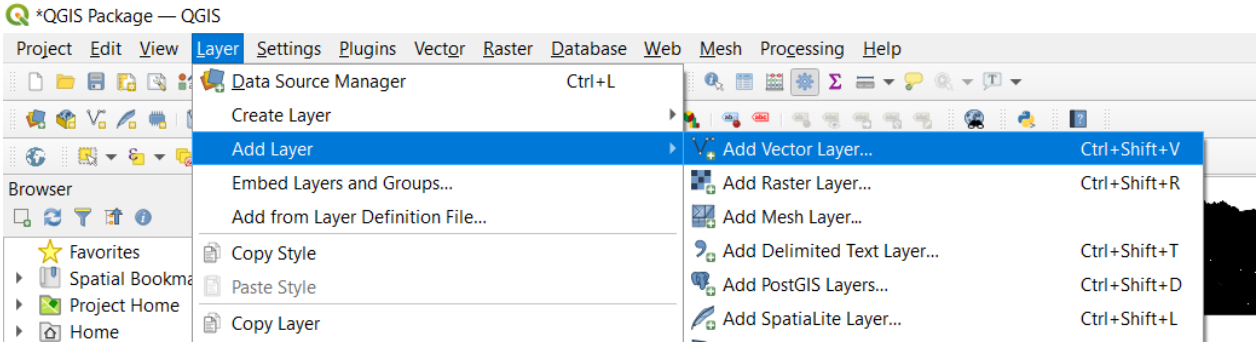
This guide outlines the steps necessary to obtain the WorldPop Top-Down population estimates using the QGIS package WorldPop Global Project Datasets, otherwise known as "**wpgpDatasets"**.

For this task, your GIS data will need to be stored as a shapefile (.shp). See Phase 1: Data Preparation, for guidance on how to collect, and if necessary, manipulate, GIS data.

4.4.1. Step 1 – Uploading your Shapefile to QGIS

First open QGIS. Looking at the tools bar, select “Layer”. Then select, “Add Layer”, and “Add Vector Layer”, as indicated in Figure 16. Locate the shapefile with the information of your administrative data. Open the folder and click on the file with the extension “.shp”.

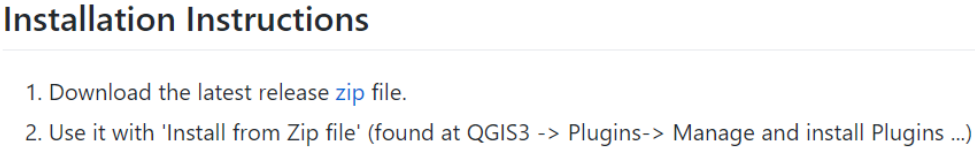
Figure 16. Example of pathway through QGIS required to access "Add Vector Layer" option.



4.4.2. Step 2 – Adding the wpgpDatasets package to QGIS

WorldPop has created detailed and thorough instructions for downloading and setting up the wpgpDatasets in QGIS as a plugin. These are available through GitHub here, or through bit.ly/3lycLRO. Complete actions indicated in points 1 and 2 under the header “Installation Instructions”, as shown in Figure 17. This website is written in English, therefore if another language is required, it is suggested to use the Google Chrome translate option mentioned in section 2.1. Purpose of the SAPIENs Guidance Document.

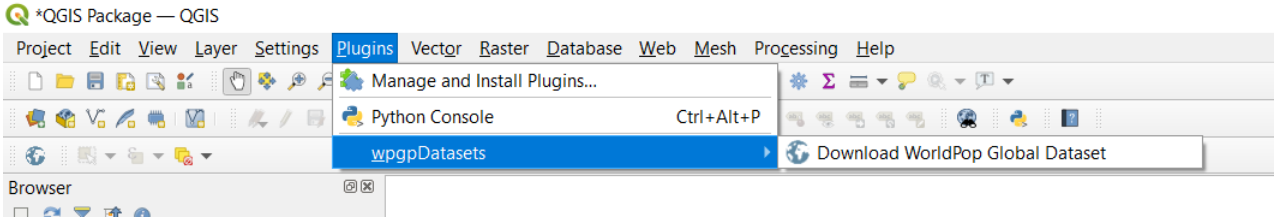
Figure 17. Screenshot of Installation Instructions of GitHub page to install wpgpDatasets package



4.4.3. Step 3 – Downloading the selected dataset using wpgpDatasets package

After adding the wpgp package as outlined by the WorldPop GitHub site, select the “Plugins” tab, then select “wpgpDatasets” and click on “Download WorldPop Global Database” (Figure 18).

Figure 18. Pathway in QGIS to access "Download WorldPop Global Dataset", otherwise known as "wpgp Datasets".

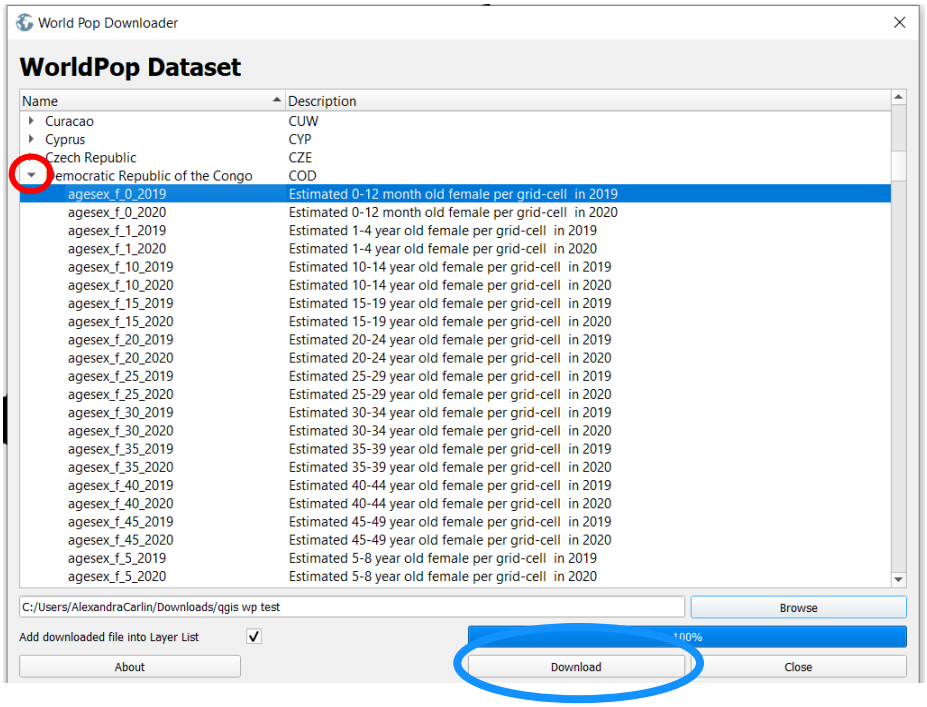


4.4.4. Step 4 – Using wpgpDatasets

Click on the country of choice using the downward pointing arrow (red circle - Figure 19). This will display the available population data produced by WorldPop as TIFF files, using the top-down modelling, for that country. The population data is available as the total population by year or populations disaggregated by sex, age, and year. For instance, the use of “agesex_f_0_2019” would be the number of females of the age from 0 to 12 months in the population, as seen in Figure 19 below. Note that you can only download one population TIFF file at a time.

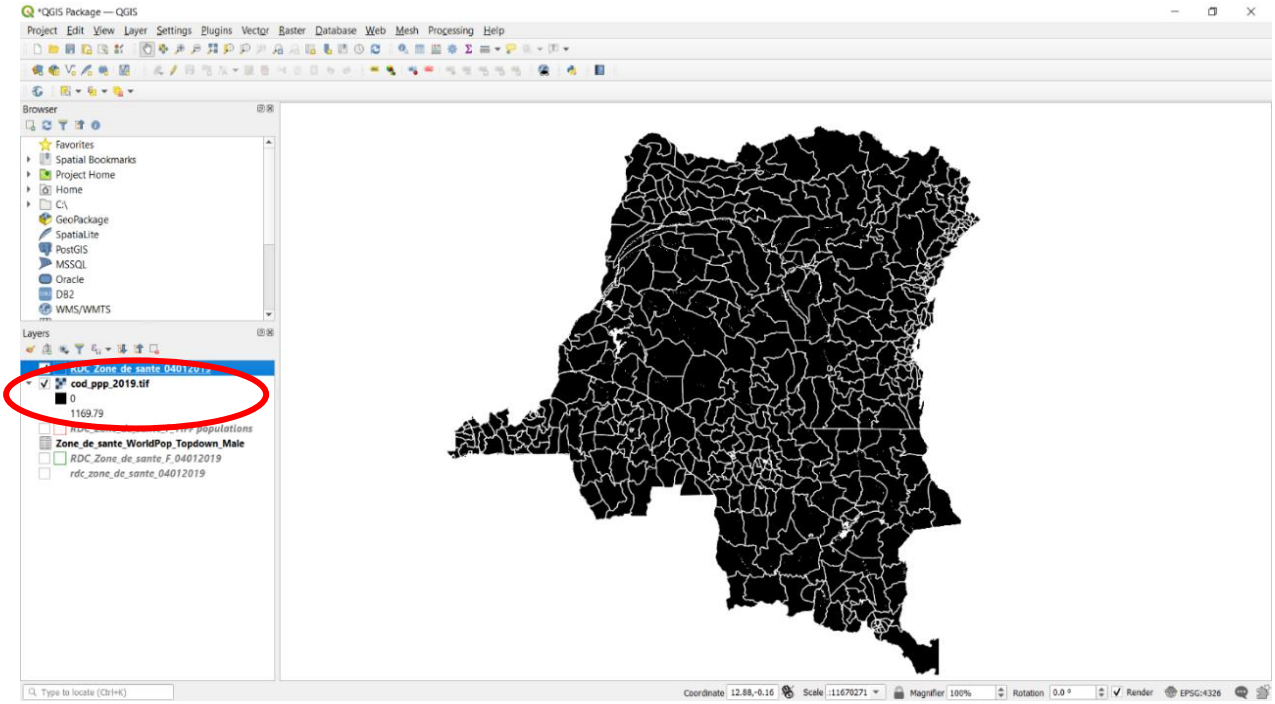
For the purpose of this exercise, we will need to click on “cod_ppp_2019”. To understand the naming convention used by WorldPop it is the 3 letter ISO3 code for the country in question, then ppp for “population per pixel”, and the year of the output. After this click on “Browse” to select where you wish to save your file and click download (blue circle - Figure 19). The progress of the download is available to be seen through the download bar as a percentage (%), in Figure 19 you can see it is 100% and totally blue.

Figure 19. World Pop Downloader Pop-up, selecting the file for total DRC population in 2019.



The “cod_ppp_2019” TIFF file will then appear as a layer in the “Layers” box on the left-hand side (red circle - Figure 20) and show as the selected map in the centre panel. Here is an example using the file named “cod_ppp_2019” to show the DRC total population estimates per grid cell during 2019.

Figure 20. Example QGIS screen with "cod_ppp_2019" as a layer in Layers box.

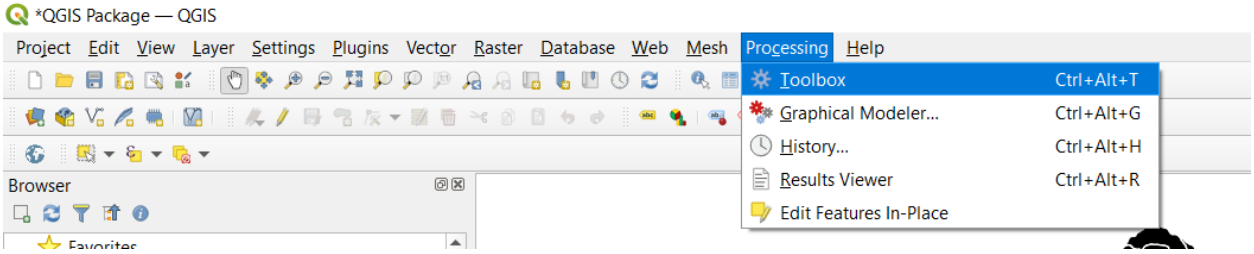


4.4.5. Step 5 – Getting administrative unit level population data.

The population data downloaded using wpgp Datasets is coded as a single layer of gridded cells covering the whole extent of the country. This kind of file is called a “raster”. Therefore, a secondary step is necessary using QGIS and the shapefile collected earlier to collect the population information to administrative unit you require.

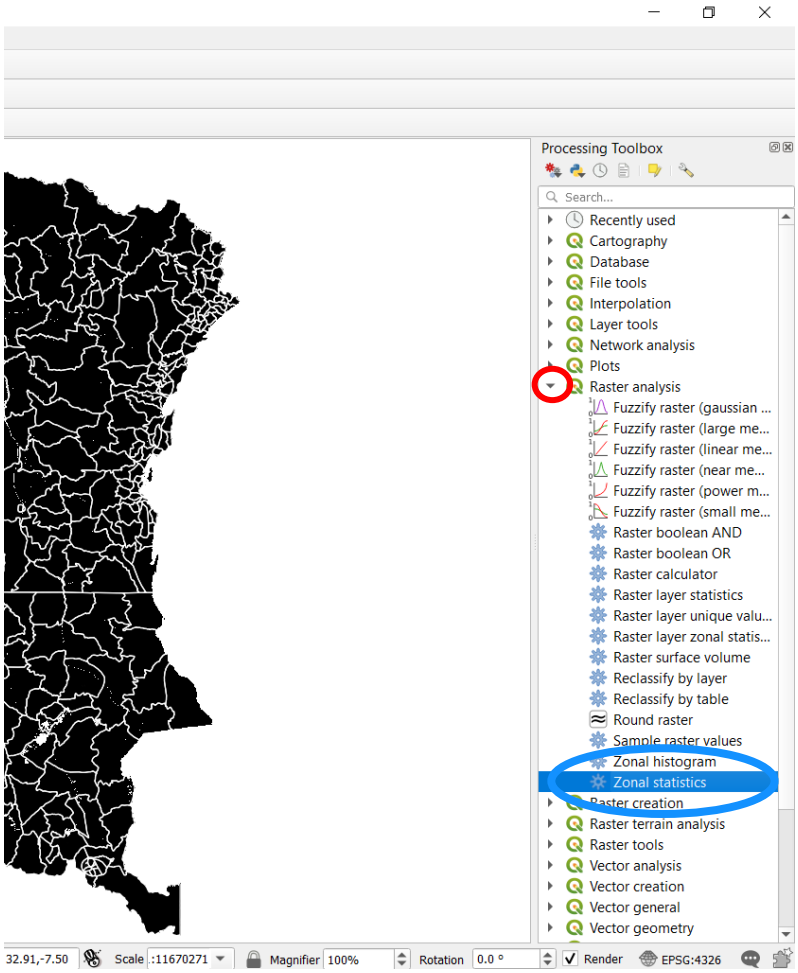
To extract this information, we need to use a processing tool called “Zonal Statistics”. To find this click on “Processing” and then “Toolbox”, as shown in Figure 21.

Figure 21. Pathway in QGIS to get to the "Toolbox" option through the "Processing" tab.



This will open the Processing Toolbox along the right-hand side as seen in Figure 22. Clicking the downward arrow for section “Raster Analysis” (red circle - Figure 22), you will find “Zonal Statistics” (blue circle - Figure 22). Click on “Zonal Statistics”.

Figure 22. Example of the Processing Toolbox box to the right-hand side of a QGIS display.



For instance, using this example, we will be extracting the population to the level of the DRC’s *zone de santé* (health zone). As you can see in the main window of Figure 22, the lines of the administrative boundaries of the zones de santé (white) overlay the raster (black) containing the total population data.

Zonal statistics will open the pop-up window shown in Figure 23. In this pop-up you will need to select the “Raster Layer” to use for the underlying calculation. Here, this is the population raster layer you have downloaded using wpgp Datasets. In this instance it is “cod_ppp_2019”.

Secondly you can select your “Raster Band”. This is an option for when a Raster file has multiple layers, and you need to select between them. In this instance, as WorldPop top-down raster files only contain one default option for this option, you can leave this option as is.

Thirdly, you will need to select your “Vector layer containing zones”, this will be the shapefile added earlier denoting the IU borders. In this case the layer name is “RDC_Zone_de_sante_04012019”. After this, you will need to name your file under “Output column prefix”. To aid data management we have suggested some standard variable names in Table 3 below, i.e., agesex (as either M or F) then an underscore, for instance 75M_.

Figure 23. Zonal Statistics Pop-up on QGIS

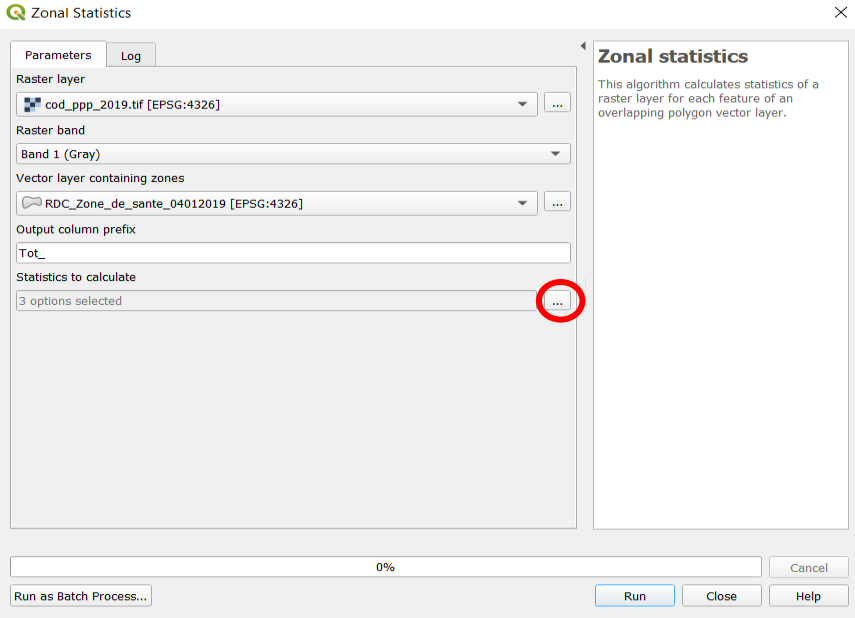
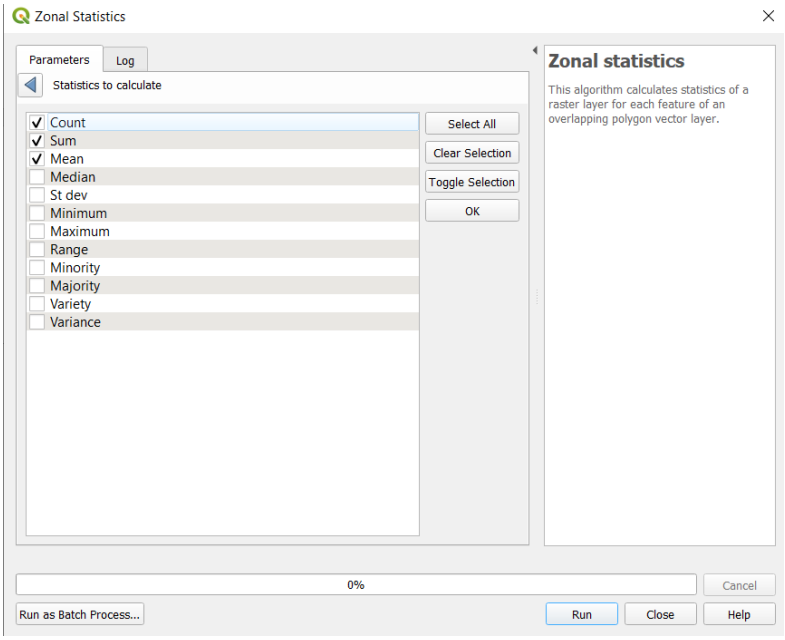


Table 3. Suggested Output column prefixes for various Raster file names

Raster file name	Output column prefix
ppp_year	Total_
agesex_f_0_year	0F_
agesex_f_45_year	45F_
agesex_m_0_year	0M_
agesex_m_45_year	45M_

After this, clicking on the "…" button (red circle - Figure 23), you will select which parameters you would like to extract from the raster file using the checklist found below in Figure 24.

Figure 24. Zonal Statistics Pop-up Parameter selection menu.



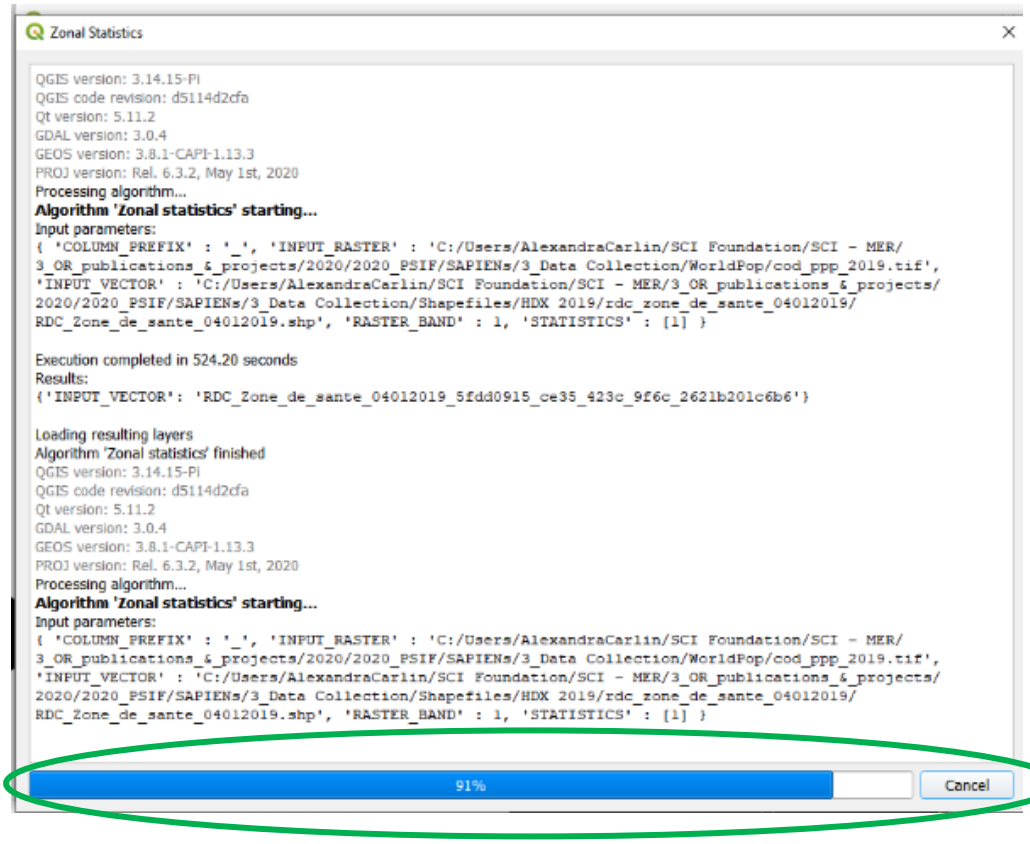
A brief explanation on these available parameters would be:

1. **Count** – The count of the number of values available in the grids of the raster file in this zone, therefore not the total. For instance, if there is a list of 1, 5, 7, 9, then the count would be 4.
2. **Sum** – The total sum of all the of values in the raster file in this zone. Thus, this would be the total estimated population for that zone.
3. **Mean** – The average value of the values in the raster grid in the zone. For example, if there is a list of 1, 5, 7, 9 as before, then the mean would be 1+5+7+9=22, 22/4 = 5.5.
4. **Median** – The median of the values in the raster grid in the zone. Therefore, if considering a list of numerical values, 1, 5, 7, 9, the median would be 6, taking the "middle" value of the list (between the numbers 5 and 7).
5. **St dev** – The standard deviation of the values in the raster grid in the zone. The Standard deviation is a method of calculating how spread out the numbers in your list are. For more about Standard Deviation, you can read here (English only) (https://bit.ly/2XIhsxg) or read here (bit.ly/3b5QIhB).
6. **Minimum** - The minimum of the values in the raster grid in the zone. Looking at the previous list, 1, 5, 7, 9, this would be 1.

7. **Maximum** - The maximum of the values in the raster grid in the zone. Looking at the previous list, 1, 5, 7, 9, this would be 9.
8. **Range** – The range (maximum – minimum) of the values in the raster grid in the zone. For the list 1, 5, 7, 9 this is 9 – 1.
9. **Minority** – The minority is the value with the fewest occurrences in the raster grid in the zone. For the list 1, 1, 4, 7, 7, 7, 9, 9, the minority is 4.
10. **Majority** – The majority is the values with the most occurrences in the raster grid in the zone. For the list 1, 1, 4, 7, 7, 7, 9, 9, the majority is 7.
11. **Variety** – The variety is the number of distinct values in the raster grid in the zone. For the list 1, 1, 4, 7, 7, 7, 9, 9, the variety is 4.
12. **Variance** – The variance is the square of the standard deviation of the values in the raster grid in the zone.

For the purpose of this investigation, you will need to select only **Sum**, as this will show us the total sum of all the of values in the raster file each IU. After this, select “Run”. The pop-up will appear similar to that is displayed in Figure 25. The download progress will be evident through the blue bar at the bottom of the pop-up via a percentage (green circle – Figure 25) .

Figure 25. Screenshot of the Zonal Statistics processing display and completion progress bar to 91%.



4.4.6. Step 6 – Repeat for all necessary agesex layers.

Repeat steps 3 to 5 again to gather further populations you may need for programme planning. For instance, to collect population data for individuals of 0-12 months for DRC 2019, you would first need to select the COD raster named “Estimated 0-12 month old females per grid cell”: “**agesex_f_0_2019**”, complete steps 3-5 again. Then you would need to gather the raster for “Estimated 0-12 month old males per grid cell”: “**agesex_m_0_2019**” and repeat steps 3-5 again. Finally, you would need to sum the male and female numbers together for the total population of 0-12 month olds in DRC for 2019.

Another example is to gather the population estimates for school-aged children (SAC) in DRC 2019, you would need to download the raster’s; “**agesex_f_0_2019**”, “**agesex_m_0_2019**”, “**agesex_f_1_2019**”, “**agesex_m_1_2019**”, “**agesex_f_5_2019**”, “**agesex_m_5_2019**”, “**agesex_f_10_2019**”, and “**agesex_m_10_2019**”. Steps 3-5 would need to be completed on each individual raster file, then all output values of sum_ would need to be added together to get the total SAC population for DRC in 2019.

4.4.7. Step 7 – Extracting population data to an .csv file.

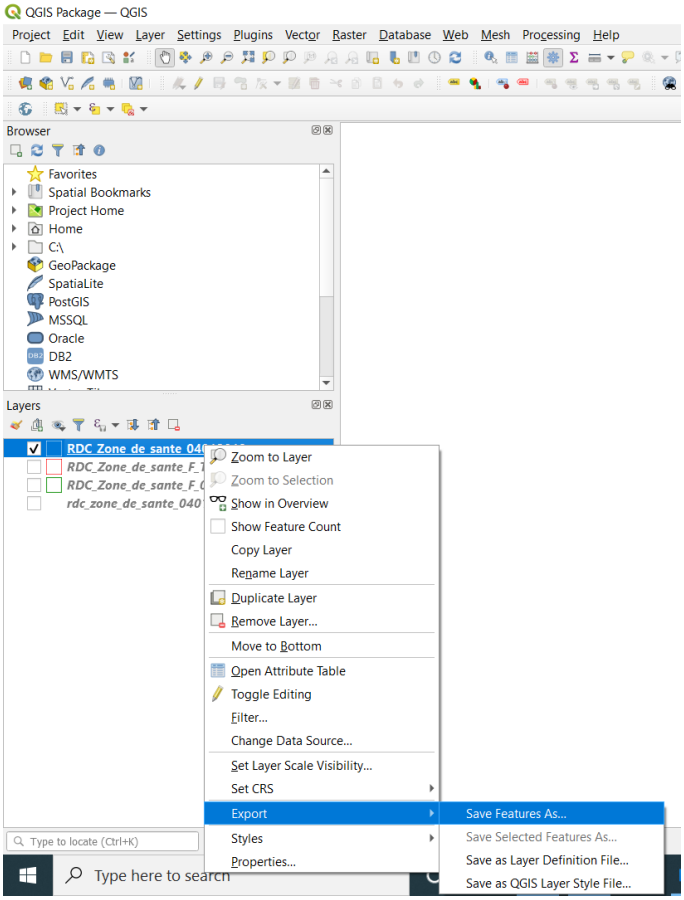
Left click on the shapefile you have been using as the “Vector layer containing zones” in the layers window of QGIS. Go to “Export” and then “Save Features As…”. The pop-up to the right-hand side will appear. The “Format” should continue to stay as “Comma Separated Value [CSV]”.

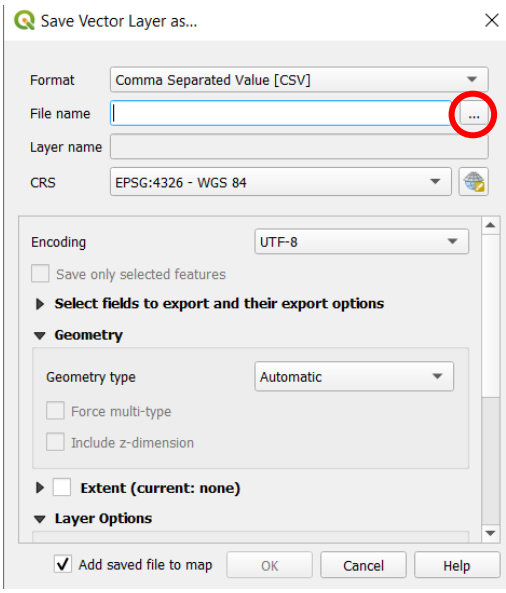
Click on the “…” next to the File name option (red circle, Figure 26) and find your chosen file location for saving, remembering to write your required “File name” into the space available, and click OK. In Table 4 below some suggested standardised file name options have been provided.

Table 4. Table of suggested standardised file names for various raster file options.

Standardised File Names
total_pop_year_admin_X_modelled 2
agesex_f_year_admin_X_modelled 2
agesex_m_year_admin_X_modelled 2

Figure 26. Pathway to "Save Features As..." option (left) and the subsequent "Save Vector Layer as..." pop-up (right).





4.5. Number of Treatments

This is defined as the number of people treated reported by your NTD Programme within the MoH after an MDA. If you are a third party, this information should be requested from the MoH. If not available through this mechanism, it can be collected within the country's JRF form on the ESPEN Portal, if available.

Note it is important to consider the relevant age groups for the specific MDA of your interest. For instance, as an SCH MDA often only targets SAC populations, the reported values will only refer to SAC. If you are considering lymphatic filariasis (LF), the target population is defined as the whole IU population[10]. Therefore, the NTD you are considering will determine the treated population for which you collect data.

It is suggested that you save this file using the name Treatments_Source_CCC_Year. Therefore, for a 2019 DRC MDA this would be Treatments_MoH_COD_2019.

[10] Guideline: Alternative mass drug administration regimens to eliminate lymphatic filariasis. Geneva: World Health Organization; 2017. Licence: CC BY-NC-SA 3.0 IGO.

4.6. Data Input into Population Review Workbook

Once you have collected and saved all the population data, the next step is to input all this information into the Population Review Workbook. Before completing this section, make sure the review workbook has been prepared as indicated in section 3.4 Preparation of SAPIENs Population Review Workbook.

Firstly, add in the IU names from the Population 1: Drug Request under the corresponding Admin level column i.e., Admin 1. Then add in the corresponding IU population figures from Population 1: Drug Request.

Next add in the population figures for the other populations you have collected for each IU in their corresponding sections: Population 2: Treatment Reporting, Modelled 1: Bottom-Up, Modelled 2: Top-Down, and Number of Treatments. However, as you are inputting these population figures record all differing administrative unit names in Sheet: 5. Admin Name Variations.

For large datasets, the VLOOKUP function on Excel could save some data input time. VLOOKUP, which stands for 'Vertical Lookup', looks up a common value across two tables and assign values from the second table to the first. If you have two tables with IU information (the first with the number of people treated the second with population data), you can use the IU name as the common variable to return the population value. More information on the VLOOKUP function is available here (English) (https://bit.ly/2Ko1AwY) or here (French) (bit.ly/3sGe6Z).

Once you have inputted all data collected, you can move onto Phase 3: Data Analysis & Visualisations.

5. Phase 3: Data Analysis & Visualisations

Within the Population Review Workbook there are various analytical and visualisation options which are designed to help display key findings and outputs from the investigation.

This section of the guidance illustrates the different analyses across the different sheets, providing examples of each table or visualisation which will be produced through the completion of the Population Review Workbook.

5.1. Sheet: 2. Data Collection

Sheet 2: Data Collection in the Population Review Workbook contains various automated data checks and analyses based on the population data you input. This next section will provide a brief explanation of these alongside explanations to further understand these analytic outputs.

5.1.1. Total of PreSAC, SAC and Adults

Within the *Total* columns of each section, conditional formatting has been set to identify when the total population entered does not match the sum of the age disaggregated population figures. Specifically, these cells will turn red filled with black text in case of a sum mismatch.

For instance, let us take the example in Table 5. In this example, there are two instances where the total population recorded in the Population source does not sum up to the total of PreSAC, SAC, and Adults. In such instances the Total population cell for that district is filled red with black text.

Table 5. Example of Conditional formatting in the Population Review Workbook, Sheet: 2. Data Collection.

Admin 3	Total	PreSAC (>5 years old)	SAC (5-14 years old)	Adults
District 1	100,000	25,000	25,000	25,000
District 2	90,000	10,000	30,000	50,000
District 3	120,000	5,000	20,000	70,500

5.1.2. Population Differences

For Population 2: Treatment Reporting, Modelled Population 1, Modelled Population 2, and Population 3: Third Party (*optional*), the spreadsheet will automatically calculate the population differences between the alternate population source, and the Population 1: Drug Request. The results are provided in the columns named “Population 1 vs …..”, where “….” stands for the comparison population source.

These calculations have been completed as both the difference in number between the two Total populations (columns “Difference (#)”) and the percentage difference between the two Total populations (columns “Difference (%)”). The “Difference (#)” is calculated by a formula which deducts the comparison population from Population 1: Drug Request. The “Difference (%)” is calculated as the value from “Difference (#)” divided by Population 1: Drug Request.

A positive value indicates that Population 1: Drug Requests is larger than the comparison source. A negative value indicates that Population 1: Drug Requests is smaller than the comparison source. This can all be visualised in Table 6 below.

Table 6. Population Differences between Population 1 and Population 2 total population column example.

Population 1 vs Population 2 (Total)	
Difference (#)	Difference (%)
-2,923	-4%
1,125	1%
-2,471	-3%

5.1.3. Reported Coverage (SAC)

Reported coverage per IU is a regularly published indicator after an MDA to assess achievement. As an example, for the template, the SAC population is used. If your NTD requires a review of total populations, i.e., LF, the formulas in these rows can be changes accordingly.

Reported coverage is calculated by dividing the number of individuals treated, the numerator, by the target population source, the denominator. In the Population Review Workbook, sheet "2. Data Collection" will provide results using all the different population sources as the denominator data which you have inputted into the template. This then highlights the difference between the reported coverage when different populations are used.

5.2. Sheet: 3. Data Analysis

There are various standard outputs which will be summarised automatically for you from the data provided along with one optional analysis. The analysis summaries provided are:

1. **Data Summary**: provides the number and percentage IUs greater than, less than, equal to and missing in comparison to Population 1: Drug Request.

Table 7. Example of Data Summary table with Population 1 vs Population 2 where country has 20 IUs.

	Population 1 vs Population 2	
Data Summary	**# of IUs**	**% of IUs**
Greater than Population 1	15	75%
Less than Population 1	4	20%
Equal to Population 1	0	0%
Missing in comparison to Population 1	1	5%

2. **Reported Coverage Achieved: SAC**: provides the number and percentage of IUs that are below 75% reported coverage, between 75% and 100% reported coverage, and >100% reported coverage for an SCH MDA for each population source.

Table 8. Example of Reported Coverage Achieved: SAC with Population 1 where country has 20 IUs.

	Population 1	
Reported Coverage Achieved: SAC	**# of IUs**	**% of IUs**
Below 75% reported coverage for SCH	7	35%
Between 75% and 100% reported coverage for SCH	9	45%
Above 100% reported coverage for SCH	4	20%

3. **Population Difference Thresholds (%)**: provides the number and percentage of IUs by percentage difference categories (e.g., between -1% and 1% difference or between +/-1% and +/-5%) when considering Population 1: Drug Request and the alternate population source.

Table 9. Example of Population Difference Thresholds (%) with Population 1 vs Population 2 where country has 100 IUs.

	Population 1 vs Population 2	
Population Difference Thresholds (%)	**# of IUs**	**% of IUs**
Population differences between +/- 0% and +/- 1%	25	25%
Population differences between +/- 1% and +/- 5%	10	10%
Population differences between +/- 5% and +/- 10%	7	7%
Population differences between +/- 10% and +/- 25%	3	3%
Population differences between +/- 25% and +/- 50%	30	30%
Population differences between +/- 50% and +/- 100%	12	12%
Population differences greater than +/- 100%	13	13%

4. **Population Difference Thresholds (#)**: provides the number and percentage of IUs by absolute difference categories (e.g., between -5,000 and +5,000 individuals difference or between +/-5,000 and +/-25,000 individuals) when considering Population 1: Drug Request and the alternate population source.

Table 10. Example of Population Difference Thresholds (# of people) with Population 1 vs Population 2.

	Population 1 vs Population 2	
Population Difference Thresholds (# of people)	**# of IUs**	**% of IUs**
Population differences under +/- 5,000 people	50	33%
Population differences between +/- 5,000 and +/- 25,000 people	22	15%
Population differences between +/- 25,000 and +/- 75,000 people	38	25%
Population differences between +/- 75,000 and +/- 100,000 people	12	8%
Population differences between +/- 100,000 and +/- 200,000 people	13	9%
Population differences greater than +/- 200,000 people	15	10%

5. **Impact on Drug Requests: SAC (*optional*)**: provides the total number of drugs which would be predicted to be required for the treatment of the SAC population in the selected areas. Secondly, it provides the percentage different of total number of drugs against Population 1: Drug Requests.

   As noted in the Excel sheet, this is the only analysis where data entry is required. As the population availability varies between sources, for this output you will need to select a specific Admin 2 where the majority of the population information for the IUs are available across all the sources. When you have selected the Admin 2 unit of your choice, you will need to enter the following formula "=SUM(cell range of SAC in population source for selected district) * 2.5". The 2.5 per child for PZQ drug requests is taken from the WHO JRSM, and it defined as the standard variable in this instance to calculate the total number of tablets required per child. However, should the standard value change, the value of 2.5 of drugs per children can be altered by the user as the formula is not locked.

Table 11. Example of the Impact on Drug Requests: SAC with Population 1 and Population 2.

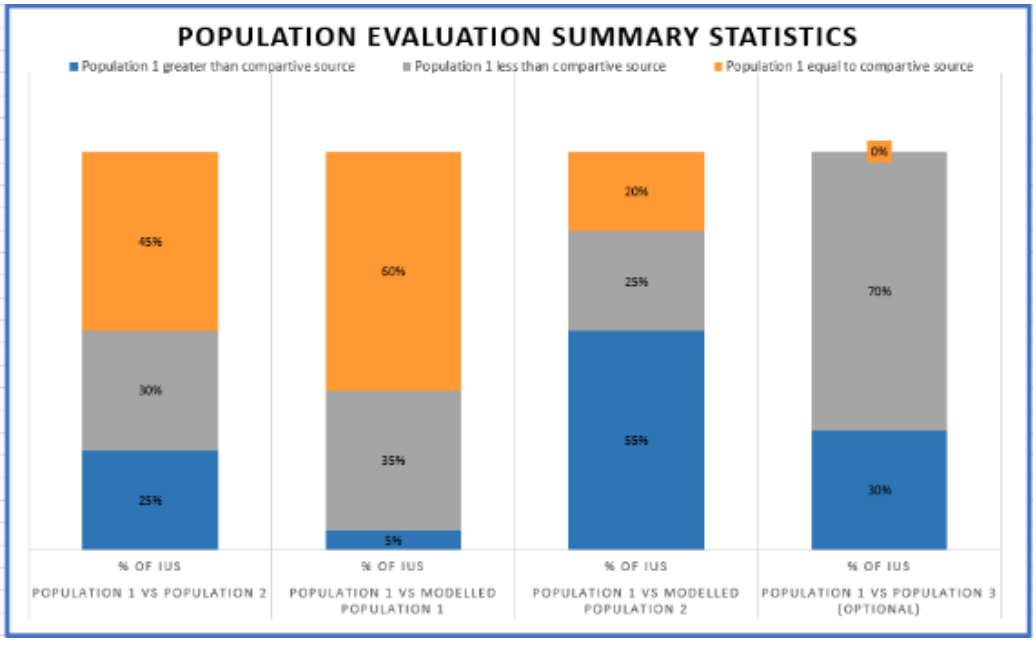
Impact on Drug Requests: SAC	Population 1	Population 2
Total # of Drugs predicted for treatment of entire SAC population where population exists	14,716,645	13,864,736
Percentage difference of total # drugs against Population 1	NA	6%

5.3. Sheet: 4. Data Visualisations

There are four standard automated data visualisations produced through the Population Review Workbook, selected for their clear representation. The visualisation should aid in highlighting key points from your investigation.

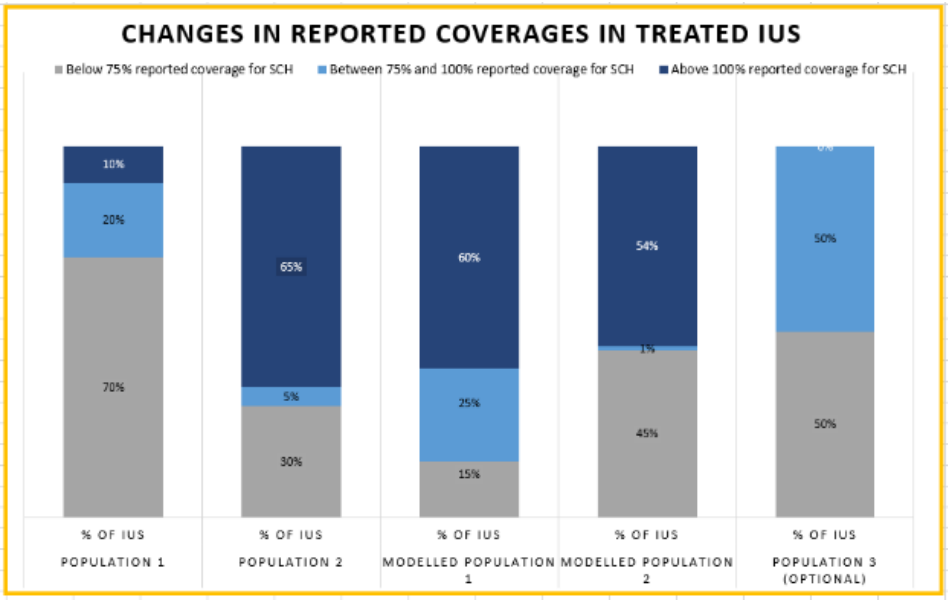
1. **Population Evaluation Summary Statistics (Figure 27):** This stacked bar chart shows the results from the *Data Summary* (Table 7), with the percentage of IUs which are greater than, less than or equal to Population 1: Drug Requests.

Figure 27. Graph from the Population Review Workbook displaying Population Evaluation Summary Statistics e.g., % greater than or less than Population 1.



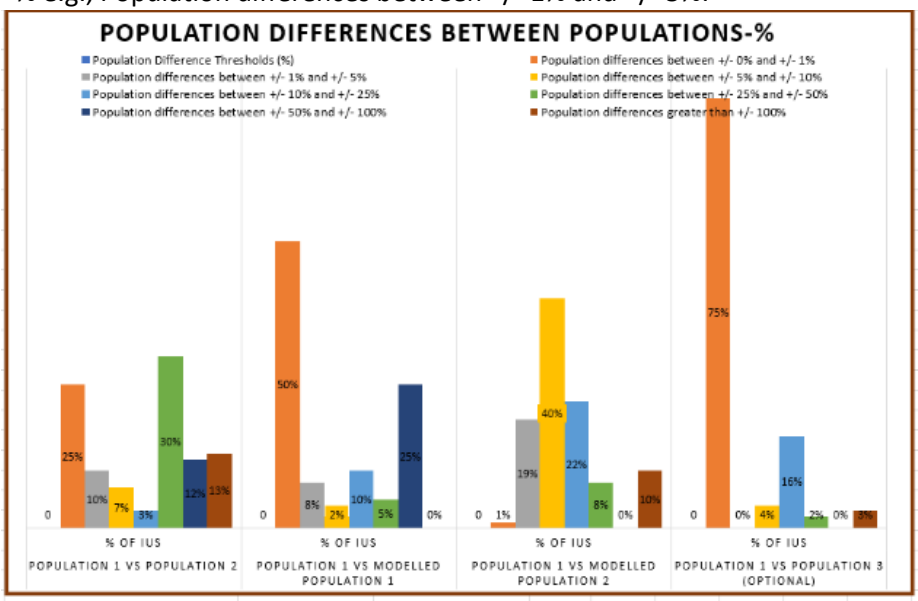
2. **Changes in Reported Coverages in Treated IUs (Figure 28):** This stacked bar chart shows the results from the *Reported Coverage Achieved: SAC* (Table 8), with the percentages of IUs across the Population sources which are below 75%, between 75% and 100%, and above 100% reported coverage for SCH.

Figure 28. Graph from the Population Review Workbook displaying Changes in Reported Coverages in Treated IUs e.g., % of IUs with cover below 75%, % of IUs with cover above 75%.



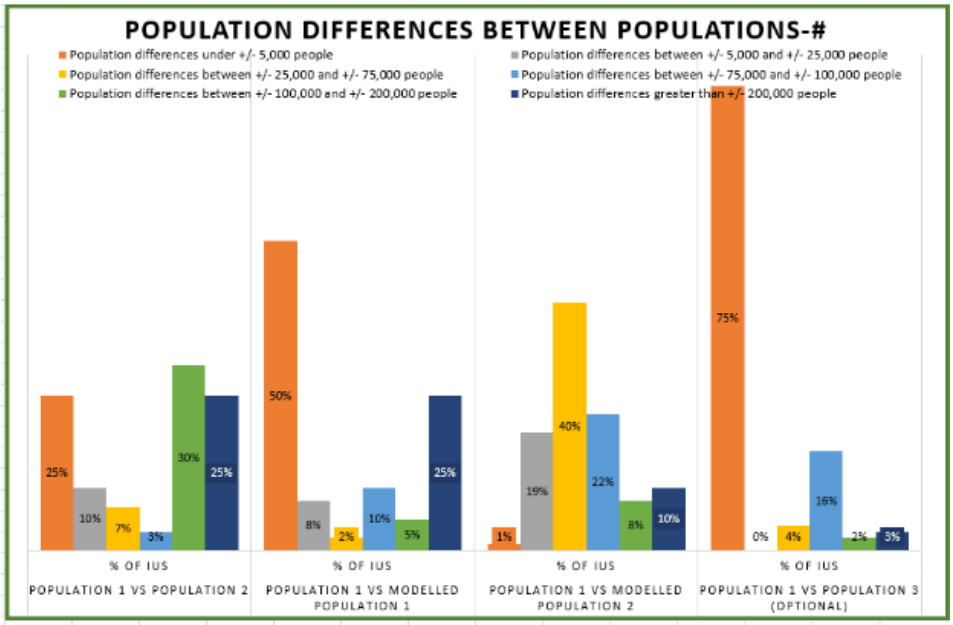
3. **Population Differences between Populations - % (Figure 29):** This bar chart shows the results from the *Population Difference Thresholds (%)* (Table 9), with the percentage of IUs within the indicated thresholds when comparing Population 1: Drug Requests to one of the alternate sources.

Figure 29. Graph from the Population Review Workbook displaying Population Differences Between Populations - % e.g., Population differences between +/- 1% and +/- 5%.



4. **Population Differences between Populations - # (Figure 30):** This bar chart shows the results from the *Population Difference Thresholds (#)* (Table 10), with the percentage of IUs within the indicated thresholds when comparing Population 1: Drug Requests to one of the alternate sources.

Figure 30. Graph from the Population Review Workbook displaying Population Differences Between Populations - % e.g., Population differences between +/- 25,000 and +/- 75,000.



6. Phase 4: Data Dissemination

Data dissemination is a complex matter which requires knowledge of audience capacity, understanding, and interest in the topic you are presenting. This guidance in this section of the document provides a general outline for a workshop; however, it should be amended where necessary to target your specific goal and/or audience/decision makers you have for your investigation results:

6.1. NTD Programme Planning and Reporting

As no 'right' population exists and possible sensitivities regarding the discussion of population data, we have suggested the following methodology to present and disseminate findings using the Population Review Workbook and discuss the pros and cons of each population source for use during NTD Programme planning and reporting. This methodology would be:

1. Identify key stakeholders involved in the use of population figures for NTD Programme planning and reporting. This may be dependent on the country; however, some examples

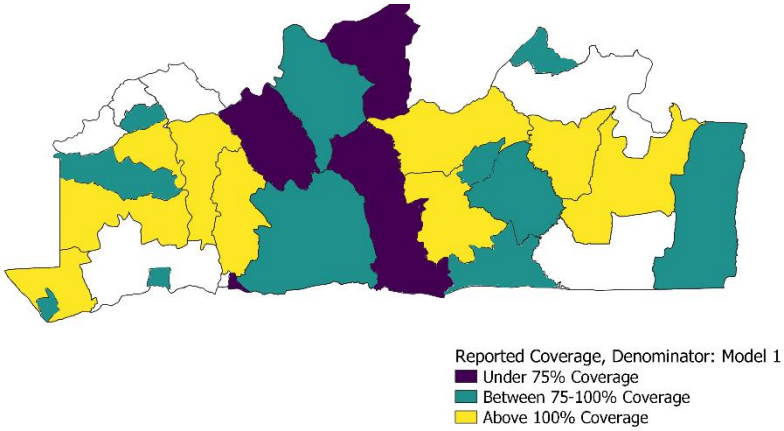
are the Lead of the MoH NTD Programme, the MoH NTD Programme Data Manager, and/or the NTD Implementing Partners.

It would be beneficial to complete a presentation to stakeholders to introduce the investigation which has been completed using the SAPIENs Guidance Document and Population Review Workbook. Specifically, this would include a brief explanation of the SAPIENs investigation process and an introduction to modelled population data to illustrate the value of the SAPIENs investigation to the NTD Programme. Information which may be useful for this presentation can be found in the Introduction section of this Guidance Document.

2. Use the standardised tables and data visualisations from the Population Review Workbook to create a presentation showing key findings from the investigation. Identify key findings from the visualisations in the Population Review Workbook, which could, for example, display the following points:
    a. What are the main differences between the population sources?
    b. How big/small are the differences between the sources?
    c. Are there any key areas/IU which have noticeable patterns in differences?

    Statements of key findings could be phrased for example as; “When comparing Population 1: Drug Requests with Population 2: Treatment Reporting for District X, there were no matching population sources. The majority of IU populations found in Population 2: Treatment Reporting were greater than Population 1: Drug Requests at 75%”.
3. Create any further visualisations which will aid the presentation of your findings. For instance, maps could be created through QGIS to show comparisons between the districts using results from the Population Review Workbook. One interesting difference for this would be when comparing reported coverage results, for instance a map similar to Figure 31 below replicated to show the reported coverage values when using each population source as the denominator. Some guidance creating basic maps is available here, or through https://bit.ly/3svbpu1. This guidance is written in English, so please use the Google Chrome translate option to translate to your required language, as necessary.

Figure 31. Map of region Kongo Central, DRC, and the reported coverages when using Modelled Population 1 as the denominator (with fabricated results).



4. When you have completed your introduction presentation to stakeholders as mentioned in point 1, a next step could be to organise a workshop to discuss results. This could be through a stand-alone workshop or incorporated into an existing country NTD planning meeting. If a workshop were preferred, the structure of said workshop could include:
    a. Presentation of current source of Population 1: NTD Planning and its limitations.
    b. Presentation of the presentation and maps created in point 2 and 3, displaying the subsequent findings from the investigation. An additional incorporation to this presentation could include an example of how modelled data or the Population Review Workbook has been used in other neighbouring or similar countries.
    c. **Group Discussion Point 1:** Looking at the findings, how could the results improve NTD Programme planning and reporting? i.e., is it possible for the NTD programme to use one population source for Drug Requests and Treatment Reporting? If not, what is needed to achieve this?
    d. **Group Discussion Point 2:** What extra support is required moving forward to increase or validate the quality of the population sources in lieu of a national census? i.e., what extra sources or partners could be involved moving forward, for instance WorldPop or UNFPA?
    e. **Group Discussion Point 3:** Are there any additional information requests for district-level MDA planning activities which are needed from areas of concern in light of the results of the Population review? For instance, in areas of with large differences between the population sources, are there data available on MDA operations to analyse the effects the use of these population sources have on MDA planning, i.e., sensitisation, drug distribution, or supply chain management? As another example, areas with concerning data may trigger the need for a district-level Data Quality Assessment (DQA) to further investigate.

f. Final thoughts and takeaways recorded by the workshop chair or lead to finalise actions moving forward.

5. Additionally, it is advised to share your findings with the country's NSO to inquire how best to proceed in the instances of significant variation in population estimates between sources.
6. Finally, it would be beneficial to share this information and decisions-made with ESPEN. This would help ensure the same population source is used across all the NTD open-source platforms.

6.2. National-level Figures

It should be noted, that sometimes changes at an NTD Planning level could facilitate larger changes at a national level. However, this would require more thought and involvement from various government bodies to achieve national-level change.

In these instances, it would be suggested to contact international stakeholders with a track record on cooperating on improving population data like UNFPA and/or WorldPop with your findings to evaluate what work is currently being done in your country and identify whether there are any instances where your investigation, or the NTD Programme, can contribute. This is to ensure there is no overlap of efforts within your country in terms of investigating population figures and encourage national ownership from the start of the investigation.

7. Appendices

7.1. Annex 1: Optional Population Sources for National-level Decision Making

As mentioned in Phase 4: Data Dissemination, there may be specific instances where you want to encourage change in national-level population figures or population figures across implementing bodies in your country. For this, an optional population source has been suggested to assess the differences in populations between Population 1: Drug Requests and a Third-Party population source used by other government bodies or NGOs.

7.1.1. Population 3: Third-Party (*optional*)

A third-party population source is defined in this context as a population source from another entity for development or humanitarian purposes. For instance, as mentioned previously, the HDX platform (found here or through https://data.humdata.org/) is a suitable platform to collect third-party information used by other humanitarian programmes.

You can search datasets using the “Search Datasets” bar to the top of the webpage, or the front page.

Figure 25. Example of the webpage for the Humanitarian Data Exchange (HDX) platform.