SurveyLab

A virtual environment for learning survey methods

User Manual

Version 4.1, September 2025

Content

1. Overview

The *SurveyLab* is conceived as a tool to integrate and complement theoretical learning of survey methods and facilitate the development of practical competencies in designing and conducting population surveys.

It consists of a virtual environment hosting a simulated human population of about 250 000 individuals distributed across the 5 Regions and 110 Towns of a hypothetical country.

The simulated population is not directly representative of any real country. However, the distributions of the variables (demographic, socioeconomic, and health-related) and their relationships are realistic and derived from the analysis of population health conducted in low and middle-income countries, and the epidemiological literature regarding relationships between risk factors and diseases.

The *SurveyLab* environment is remotely accessible through a web interface. It allows users to design their survey projects, sample the population according to their plan, and collect the data of interest that they can then analyse and interpret.

2. Access

The *SurveyLab* is accessible with any modern internet browser supporting **HTML5** and **JavaScript**. There are no technical limitations to accessing it from mobile devices, and the interface will adapt to any screen size. However, activities such as manual sampling, which require exploring maps of the environment, are best conducted on large screens, and a full-HD resolution (**1920x1080 pixels**) is recommended.

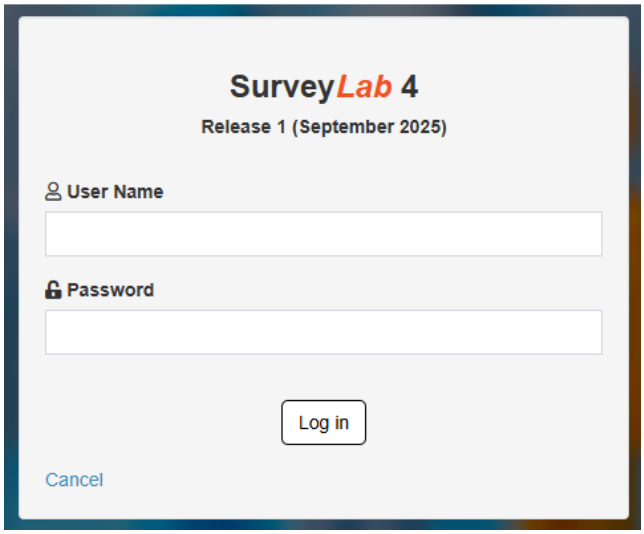
The home page shows the image depicted below.

Clicking on the button

will show a page with an overview of the *SurveyLab*.

Clicking on the button

Visualises the login screen

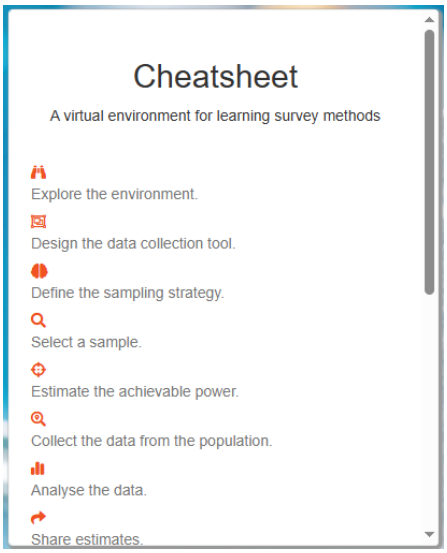


Both User Name and Password are **case-sensitive**.

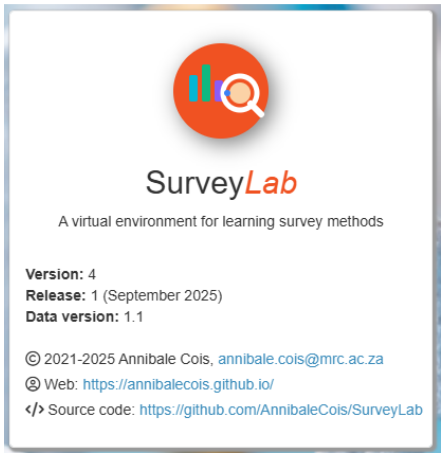
The **Cancel** link sends back to the initial screen, while pressing the Log in button Gives access to the *SurveyLab* main interface:

At the top left corner of the interface, there are three clickable items.

Item ① (Help) visualises a synthetic summary of the functioning of the various elements of the interface:

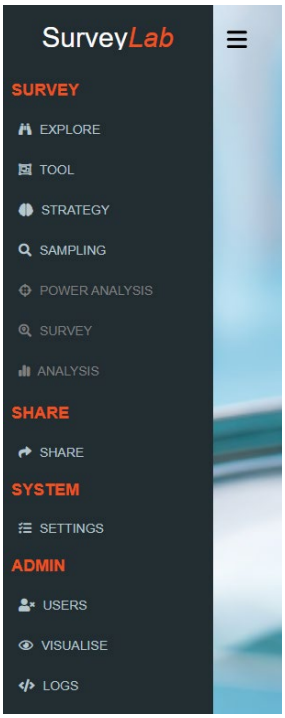


Item ② (About) visualises the Version of the system and the dataset, the copyright notice and the links to the source code and the developer's home page.



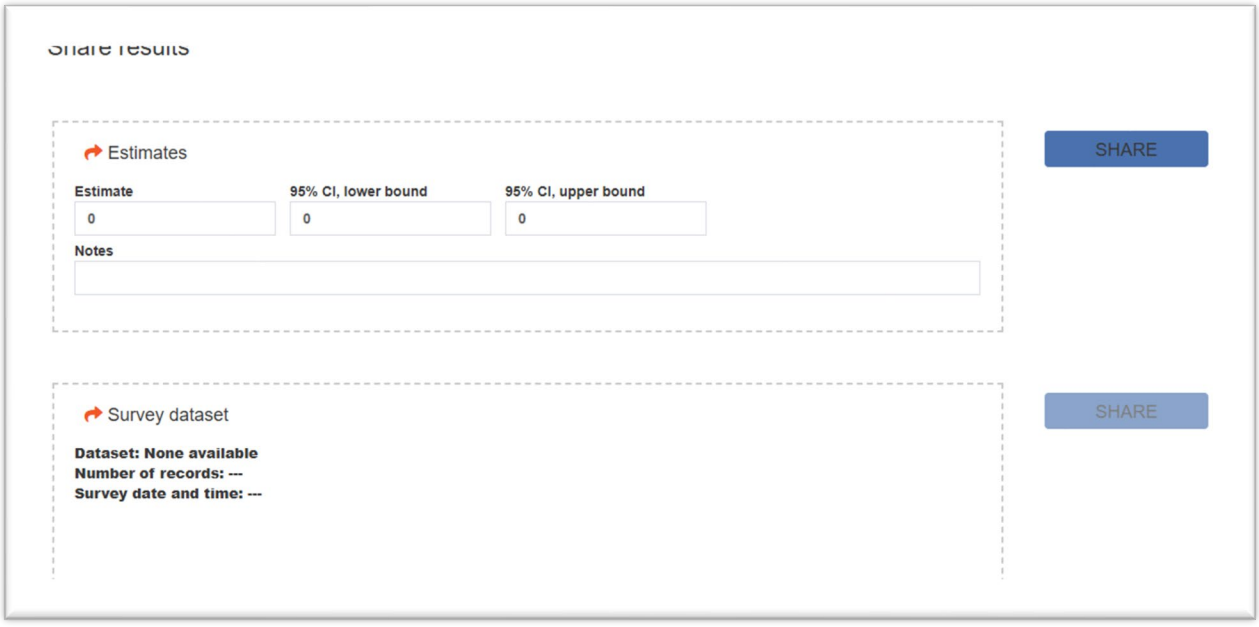
Item ③ (👤) opens a sub-menu, where the user can change the password or log out:

On the left of the screen, the main menu includes four sections: **SURVEY**, **SHARE**, **SYSTEM** and **ADMIN**:



The items in the **SURVEY** section open a set of pages dedicated to the different steps necessary to conduct a survey, which are described in the next chapter of this manual.

The only item in the **SHARE** section opens a page which allows the user to upload estimates of whole results datasets to the SurveyLab server:



Users with administrative rights (Administrators) can collate and visualise estimates uploaded by multiple users (menu items **VISUALISE** in the **ADMIN** section) and access the uploaded results datasets for further analyses.

The section **SYSTEM** (menu item **SETTINGS**) allows Administrators to change the default settings of the *SurveyLab*, such as the 'behaviour' of the virtual population members.

The section **ADMIN** allows Administrators to add and delete users and to assign/change levels of access (menu item **USERS**), and to visualise server logs (menu item **LOGS**).

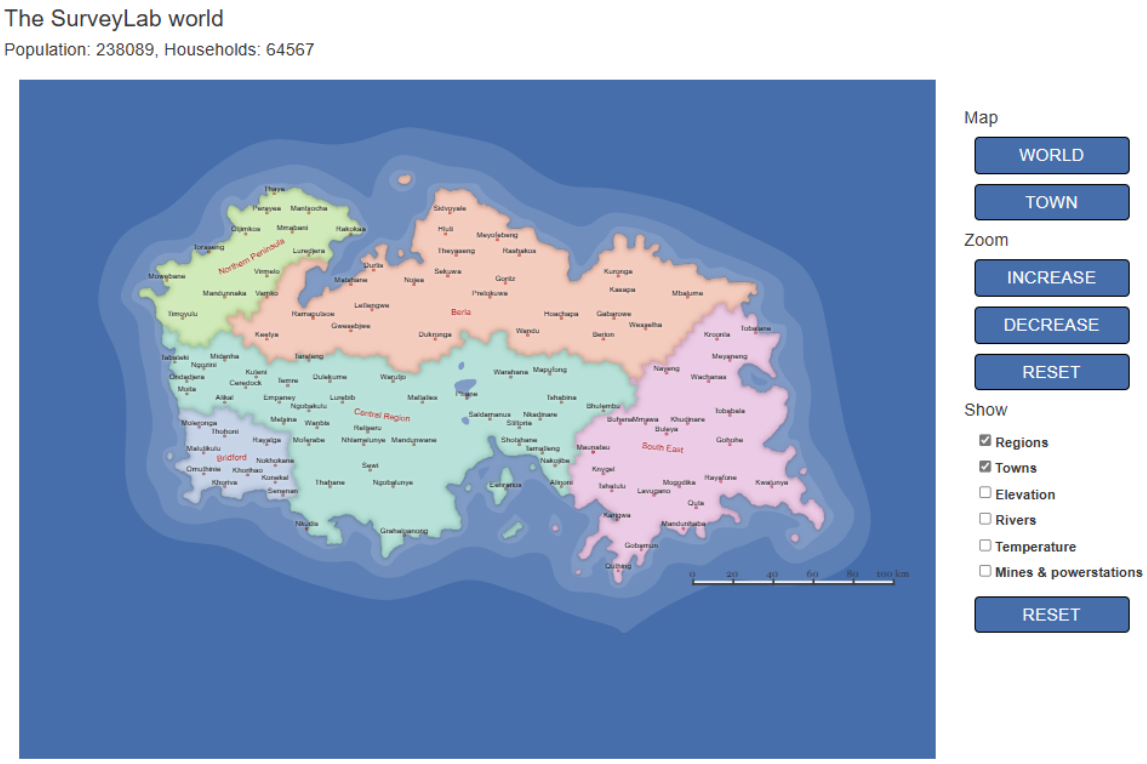
The accordion icon on the top right corner (☰) allows for collapsing the menu to save space for the main section on the right. The item names will disappear, but the icons will remain visible and usable for navigation.

3. Conducting a virtual survey

The SURVEY section of the main menu includes 7 items.

3.1. Explore

The **Explore** page presents a dynamic map of the *SurveyLab* world.



You can navigate the map with the buttons on the right:

Map

WORLD	Visualise the world map
TOWN	Opens a dropdown menu to visualise a specific town

Zoom

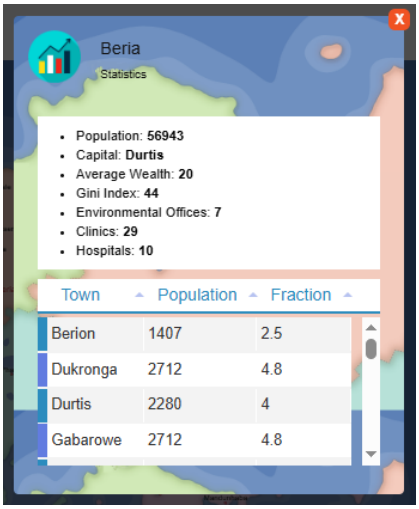
INCREASE	Increase the size of the map
DECREASE	Decreases the size of the map
RESET	Reset the size of the map to the default value

Show

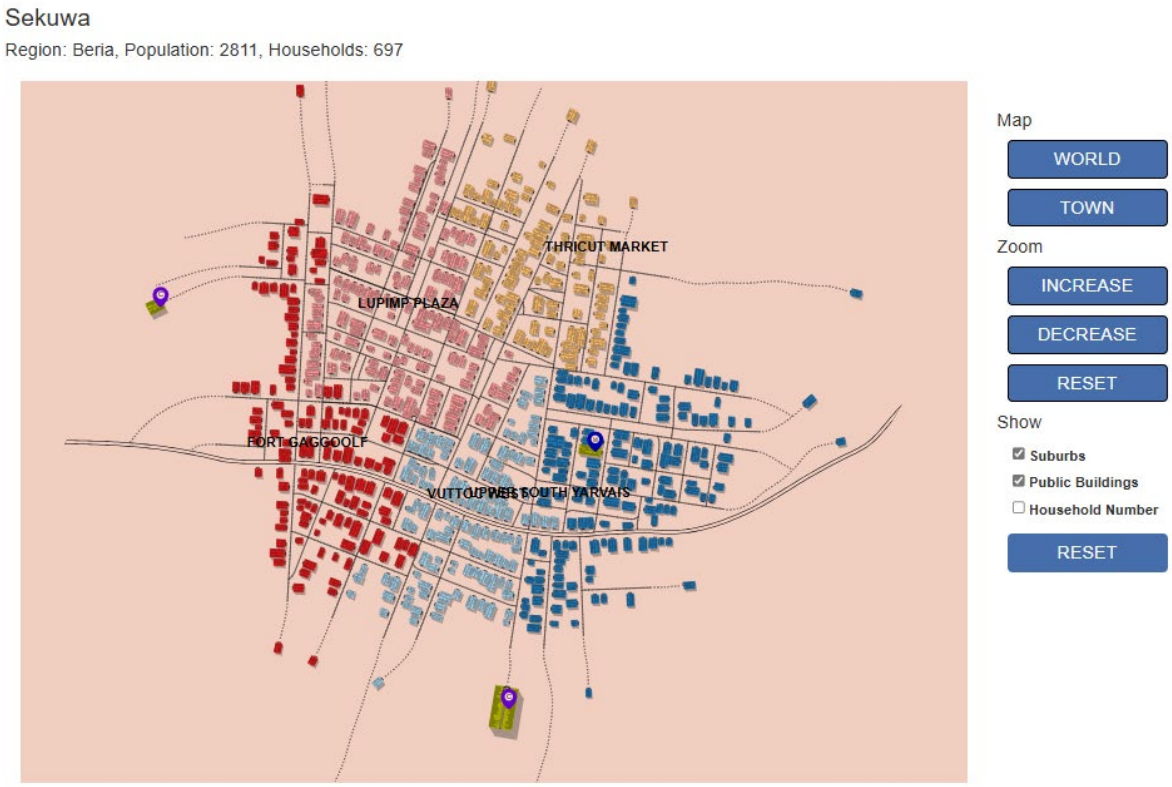
Regions Towns Elevation Rivers Temperature Mines & powerstations	Visualise/Hide various elements of the map: Regions, location of towns, altitudes of the various areas, rivers and lakes, temperature zones, location of mines and coal power stations.
RESET	Resets to the standard view (only regions and towns visible)

The interface also supports mouse operations and touch screens, and scrolling and zooming of the map can be also obtained using these devices.

Clicking on a **region name** visualises a pop-up window with information on the region, including sociodemographic indicators and the list of the various settlements (towns).



Clicking on a **town name** opens a new page relative to the chosen town:



The buttons and checkboxes on the right-hand side work as on the previous page and allow for increasing/decreasing the map size and visualising/hiding different types of information.

Clicking on the symbols below (**Public Buildings**) makes the system visualise a series of pop-up windows providing information on the Town (**Municipal Offices**, one per town), on the **Clinics/hospitals** (none, one or more per town) and about the environment (**Environmental Health Offices**, present only in the capital town of each region and a few other.)

The information provides a basis for planning the survey.

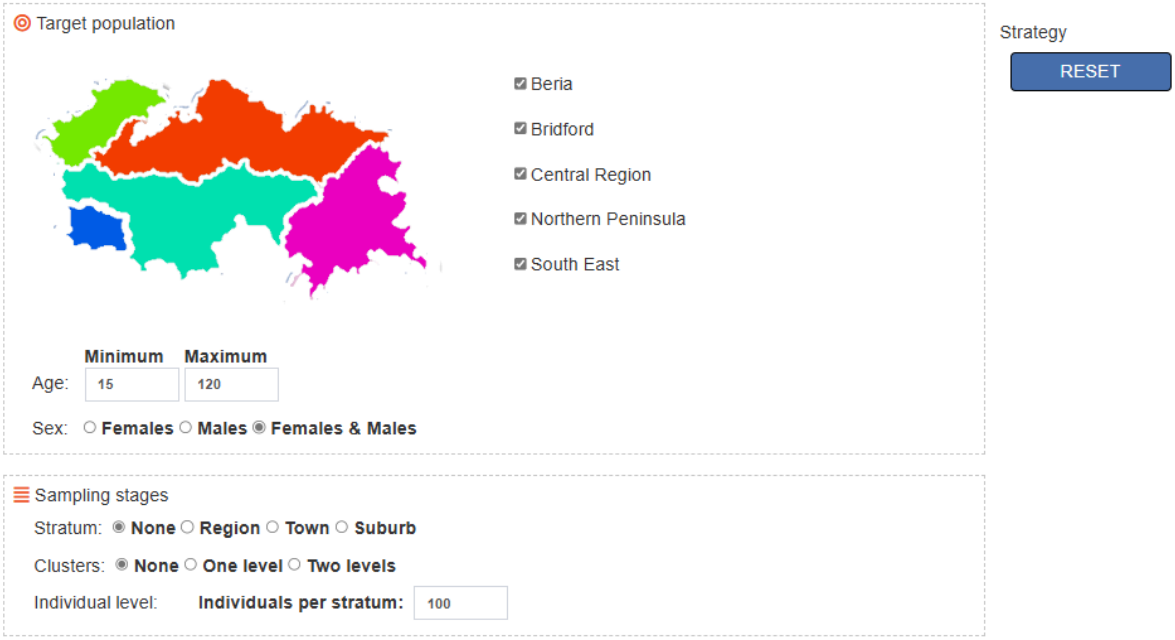
3.2. Tool

The **Tool** page allows for designing a data collection tool for the survey. The tools can include both questionnaire items and measurements and can be built by dragging the relevant questions/measurements from the lists on the right and dropping them on the left section (**Current tool**).

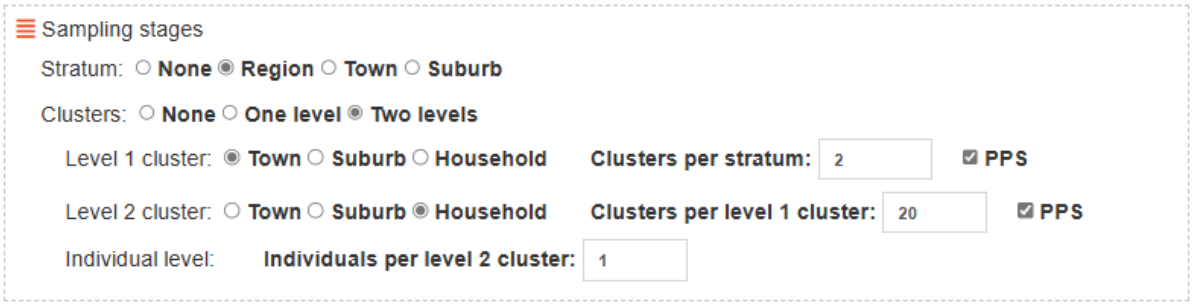
The (RESET) button empties the Current tools section and resets the lists of available items to the default status.

3.3. Strategy

The **Strategy** page allows for defining a **random sampling strategy** that the *SurveyLab* will apply to select the sample.



At the top of the page, users must define the **Target Population** by choosing the geographic scope (either the whole *SurveyLab* or specific Regions) by checking/unchecking the checkboxes, and the desired **age range** and **sex** of the potential respondents.
The **Sampling stages** section requires information regarding **stratification** (none, by Region, by Town, by Suburb), clustering (none, one level, two levels) and **sample sizes** at the various levels.
Depending on the choices, different type of information need to be provided. The most complex case is when the chosen strategy is two-level stratified random sampling, where information needs to be provided on stratification, clustering variables and sample size at each level, type of cluster selection (proportional to size or not).
An example of this case is shown below:



If the user does not make any selection, the system uses a default sampling strategy with these characteristics (**simple random sampling** of size 100):

- **Target population**: whole surveylab, both sexes, age range 15+
- **Stratification**: None
- **Clusters**: None
- **Sample size**: 100

Note: it is the user's responsibility to ensure that the sampling parameters are meaningful. The strategy page will accept any set of parameters. Only at the sampling stage (page **Sampling**, described in the next section), when the actual selection of the respondents takes place, an incongruent request (i.e. requesting to select a number of clusters greater than the number of clusters in the stratum) will cause the impossibility of carrying out the procedure as requested.

3.4. Sampling

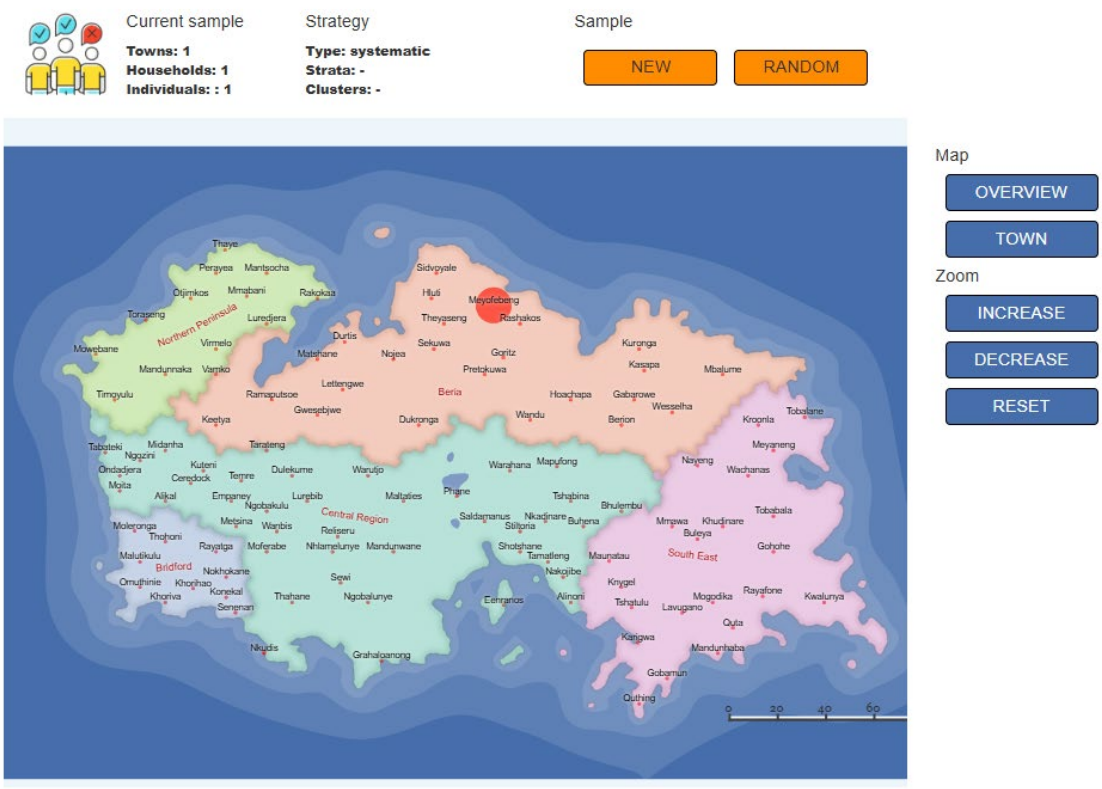
The **Sampling** page allows for selecting the individuals to whom the data collection tool will be administered during the survey phase. Two types of sample selection are possible.

3.4.1. Systematic (manual) sampling

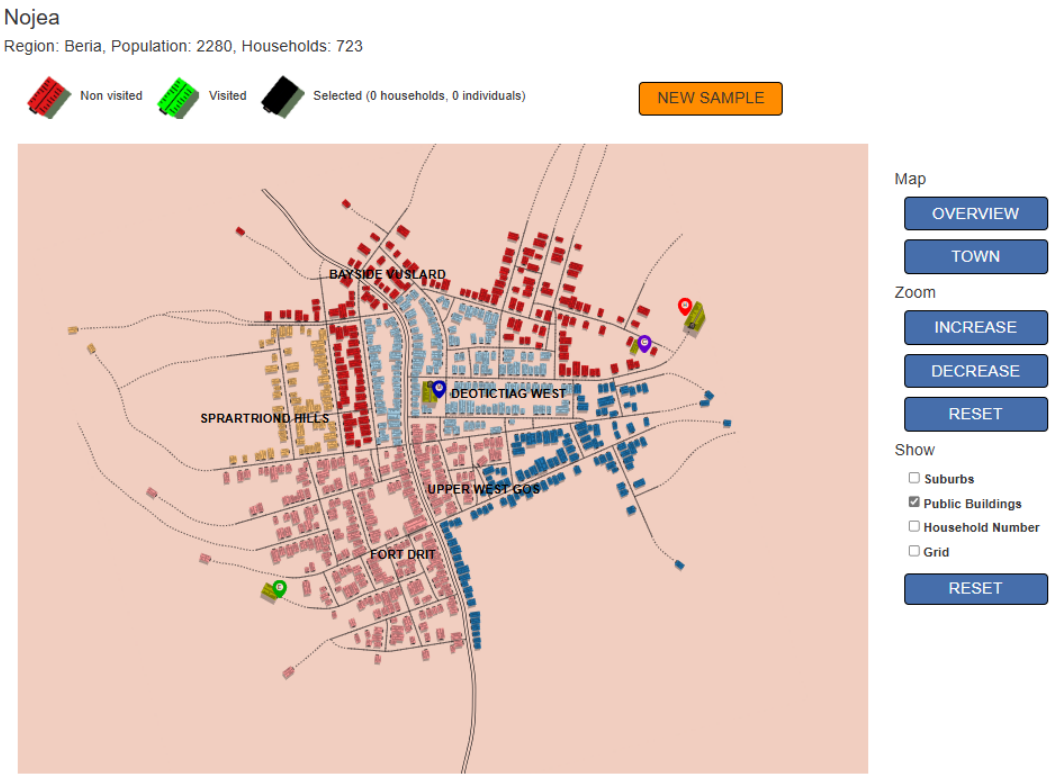
With the **systematic sampling strategy**, the user will conduct a manual selection of the individuals to be included in the sample. This procedure includes

1. Selection of **towns**;
2. Selection of **households** within a town;
3. Selection of **individuals** within a household.

The Sampling page (see figure below) looks and works similarly to the explore page:



In this case, however, clicking on a town name (or using the TOWN button on the right) leads to a page where it is possible to select households (and individuals within households) to be part of the sample

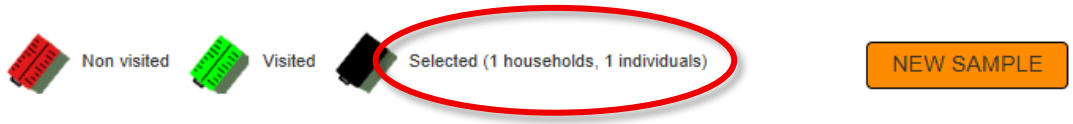


The households are selected by clicking on a dwelling: the dwelling changes colour to **green**, the household is added to the sample, and the members are listed on a pop-up window:

The household members are identified by an icon, which provides basic demographic information, as shown in the table

 Kewhea population group 0-9 years Male, female	 Ofrijan population group 0-9 years Male, female
 Kewhea population group 10-19 years Male, female	 Ofrijan population group 10-19 years Male, female
 Kewhea population group 20-39 years Male, female	 Ofrijan population group 20-39 years Male, female
 Kewhea population group 40-64 years Male, female	 Ofrijan population group 40-64 years Male, female
 Kewhea population group 65+ years Male, female	 Ofrijan population group 65+ years Male, female

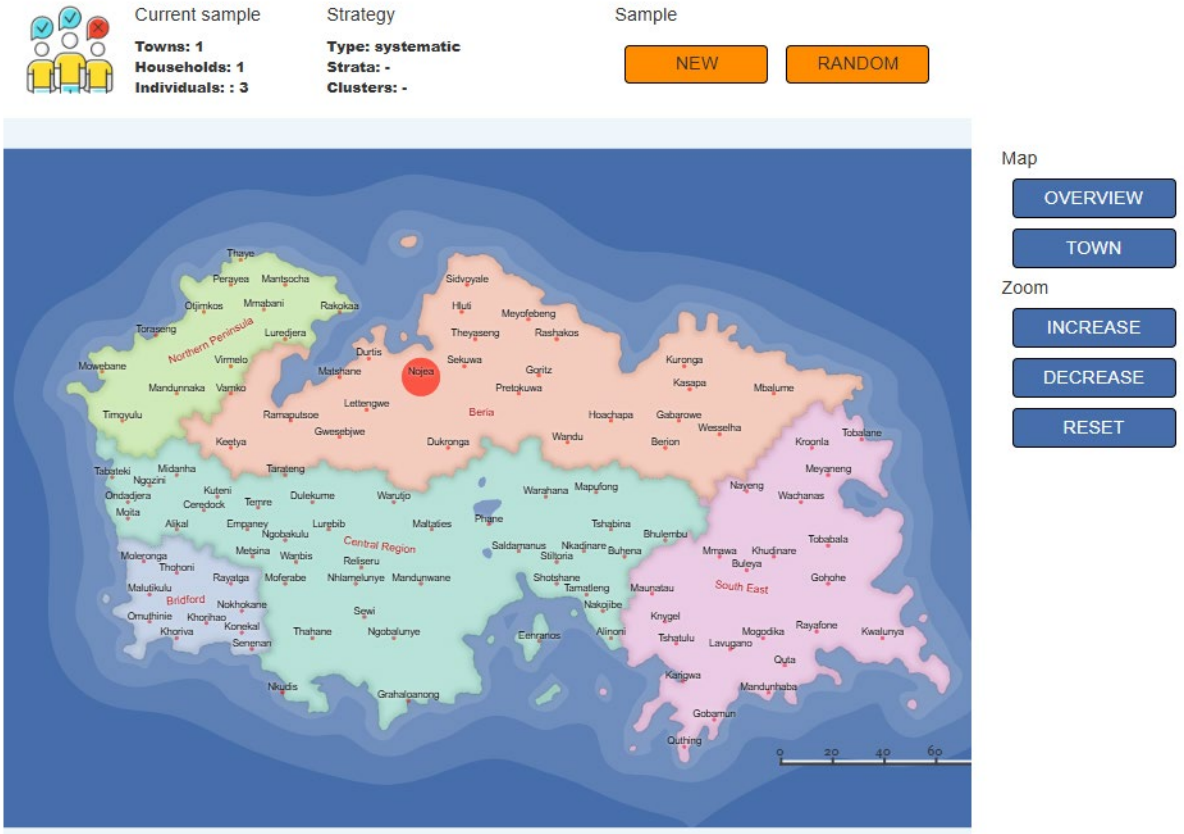
Clicking on an icon selects the individual in the sample. Clicking again deselects him/her. Pressing the SELECT button add the household (with the selected individuals) to the sample. The dwelling colours changes to **black**. The countes at the top of the page indicating the number of households and individuals selected in the town are updated accordingly.



Pressing CANCEL cancels the operation. The household is not selected, but the dwelling stays green to indicate a previous visit.

This procedure can be repeated many times to select multiple households in a town and to add multiple towns to the overall sample.

The OVERVIEW button visualises a SurveyLab map with the sampled Towns highlighted as in the figure below:

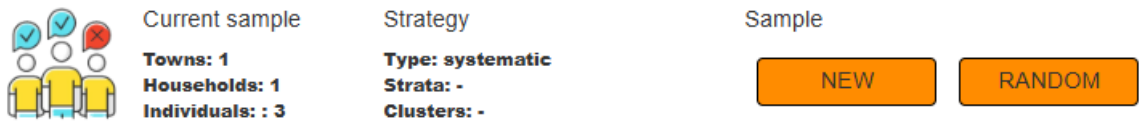


At the top, the page provides a summary of the sample (total number of towns, households and individuals selected) and the type of strategy in place ("Systematic" for the manual sampling procedure).

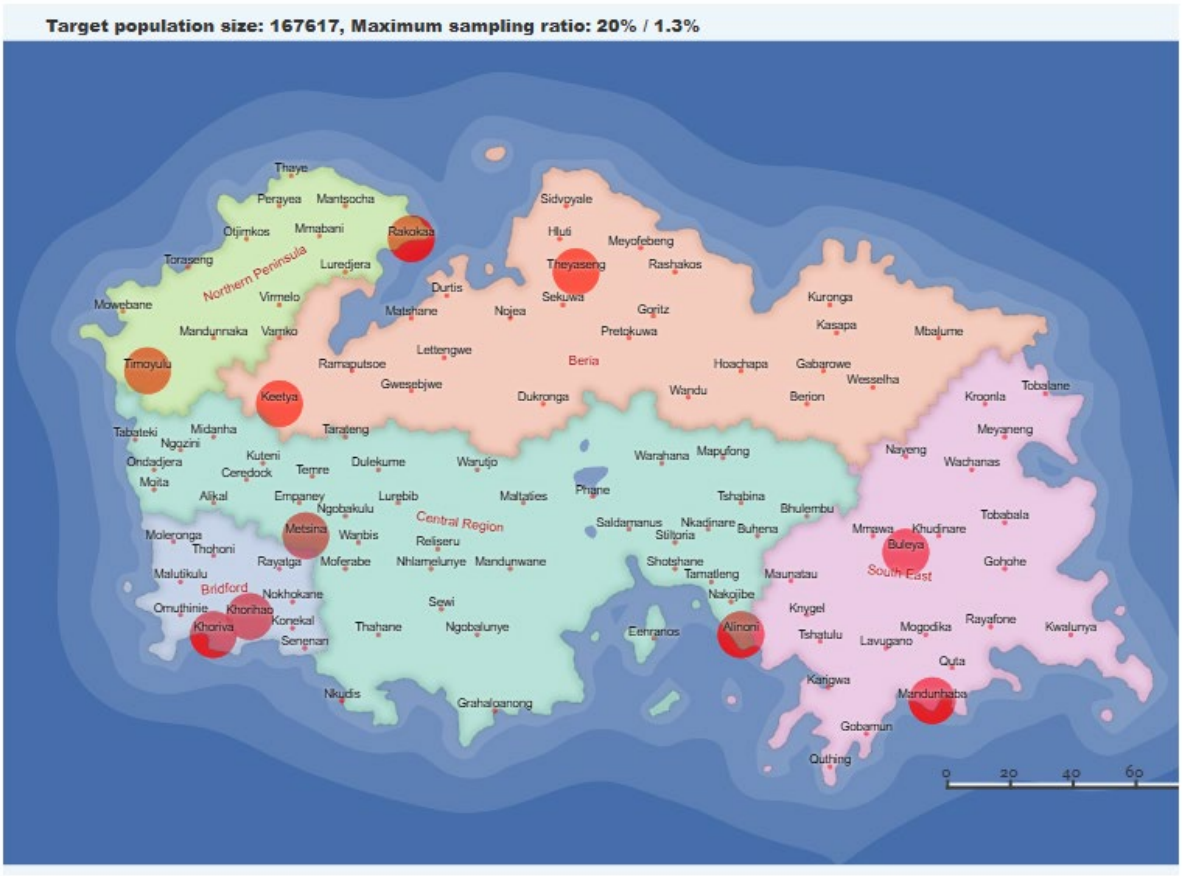
The button NEW deletes the whole sample and resets the relevant counters.

3.4.2. Random sampling

With the **random sampling strategy**, the system will randomly sample the population **according to the strategy defined in the Strategy page**. The procedure is automatic and it is triggered by pressing the RANDOM button in the overview:

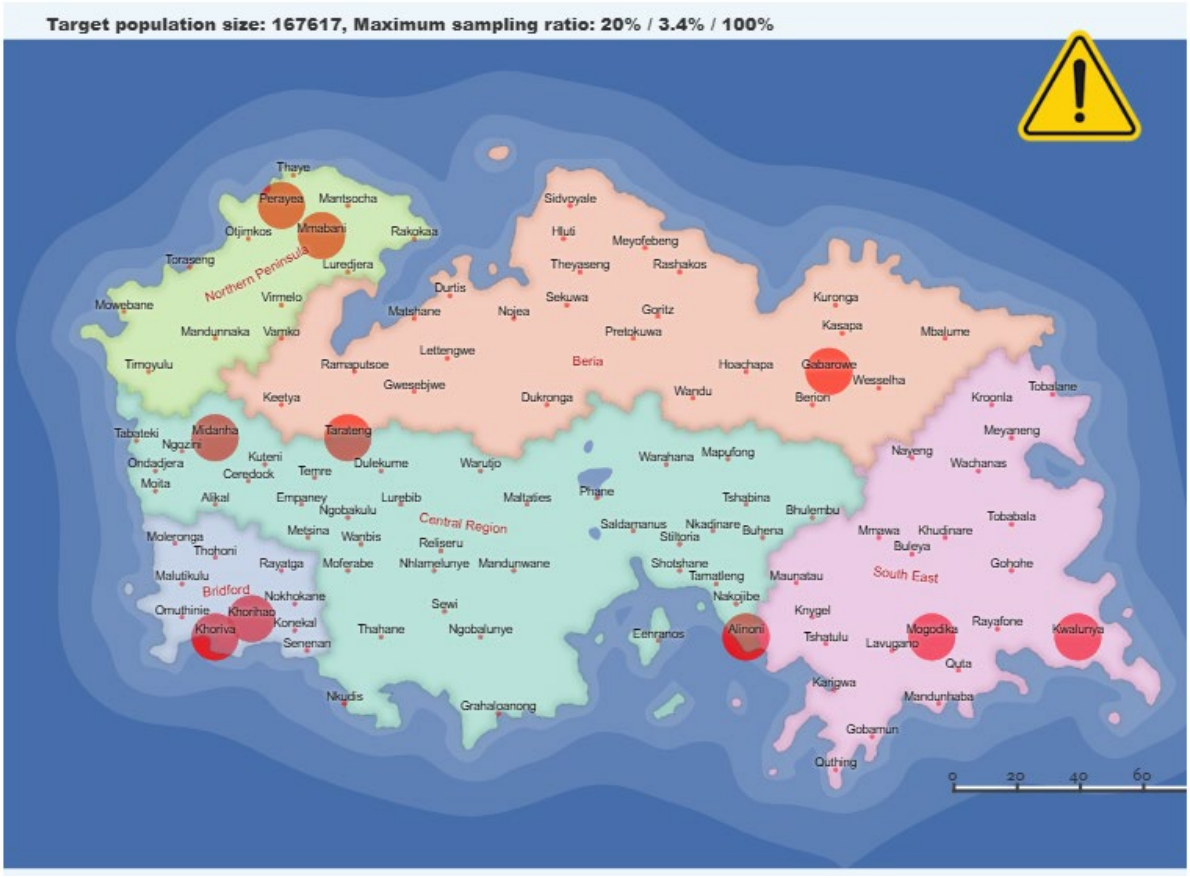


A summary of the process (which can last several seconds, depending on the sample size and the chosen strategy) is shown on the same page:

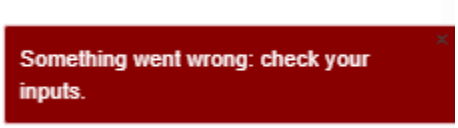


At the top of the map, the system shows the total **size of the target population** and the maximum **sampling fraction** per level.

When the sampling reaches saturation at some level (i.e. when the user requested more individuals than are actually in the population), the system still proceeds to the sampling (by including all available individuals in the sample), and notifies the user with a "warning" sign overlapping the map, as in the figure below:



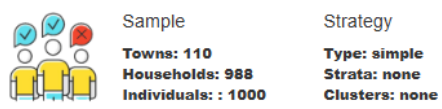
When the incongruencies make the whole sampling strategy break (e.g. indicating the same variable at different levels), the sampling cannot proceed with the chosen strategy. The system does not draw the sample and notifies the user with an error message at the bottom right corner of the page:

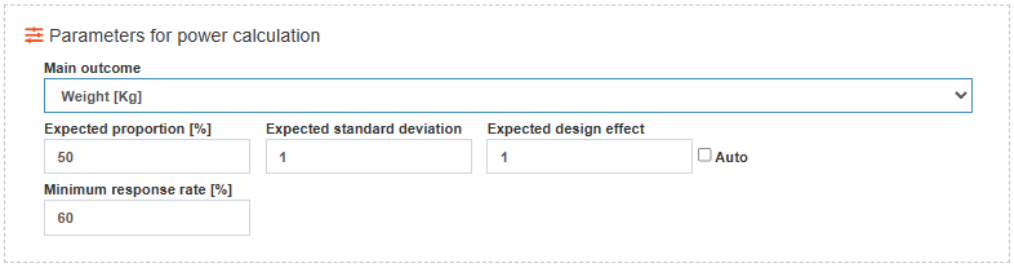


1.1. Power analysis

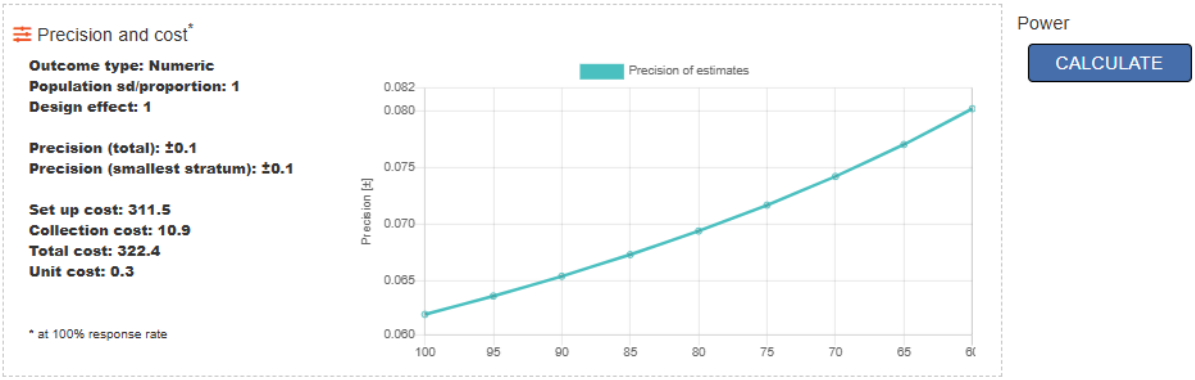
In the **Power analysis** page, the user can conduct a preliminary analysis of the precision that can be obtained with the chosen sampling strategy.

The user chooses the outcome for which the results need to be calculated, and the basic parameters upon which the calculation will be based (expected proportion/standard deviation, design effect and minimum response rate). Alternatively, the system can calculate the parameters from the underlying population, when the user checks the "Auto" checkbox.





The **Precision and cost** panel will show the results of the calculation, which include an estimate of the **total survey cost** and **unit cost**:



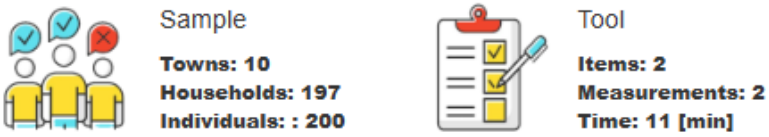
Results are recalculated either by pressing the CALCULATE button or automatically when any of the inputs change.

The system estimates administration times and costs using an algorithm that attributes different weights to different types of items in the data collection tool (e.g. it attributes longer administration times and costs to anthropometric measurements compared to questions). The calculations take into account increases in setup costs, when a survey spans across multiple towns, and the costs within each town based on the total distance a fieldworker has to walk/drive to visit each selected household.

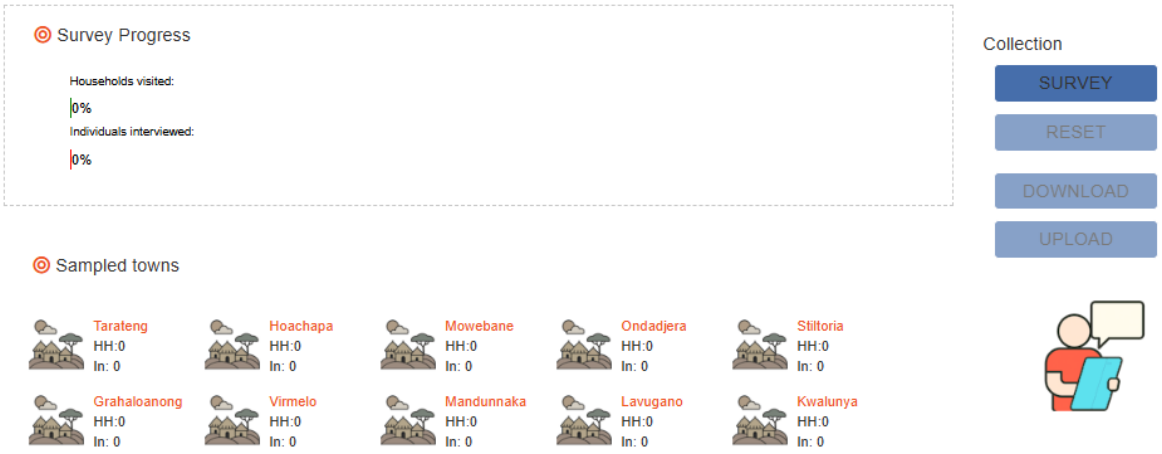
1.2. Survey

When a valid sample has been selected and a data collection tool has been defined in the respective pages, the **Survey** page allows users to start a (virtual) **data collection**.

The page shows a summary of the current sample and data collection tool:



And a graphical list of the sampled towns:



The survey can be conducted town by town, by dragging the data collector to the respective town:

or for all towns at the same time, by pressing the SURVEY button.

In both cases, the data collection tool is 'administered' to the sampled individuals and the results become available for download (DOWNLOAD) and – if enabled by the Administrator – for upload to the server (UPLOAD).

The downloaded file is in Microsoft™ Excel© format and consists of two sheets. The first sheet ("Responses") includes the individual responses, one row per individual, as in the example shown in the figure:

① IID	② HID	③ HR	④ IR	⑤ REGION	TOWN	SUBURB	SEX	AGE	WEIGHT	DISEASE_COPD	⑥ Probs
H_109_10	H_109_10	Consent	Consent	South East	Gohohe	Sphaffool	Male	42	56.8	No	8.485E-07
H_109_17	H_109_17	Consent	Consent	South East	Gohohe	East Blihia	Male	53	60.5	Yes	8.485E-07
H_109_19	H_109_19	Consent	Consent	South East	Gohohe	East Blihia	Female	53	97.1	No	8.485E-07
H_109_20	H_109_20	Consent	Consent	South East	Gohohe	East Blihia	Male	65	62.4	Yes	8.485E-07
H_109_23	H_109_23	Consent	Consent	South East	Gohohe	East Blihia	Female	35	40.6	No	8.485E-07
H_109_27	H_109_27	Consent	Consent	South East	Gohohe	East Blihia	Female	33	70.5	No	8.485E-07
H_109_29	H_109_29	Consent	Consent	South East	Gohohe	Wullaind '	Male	29	64.5	No	8.485E-07
H_109_34	H_109_34	Consent	Consent	South East	Gohohe	Wullaind '	Female	41	44.8	No	8.485E-07
H_109_39	H_109_39	Consent	Refusal	South East	Gohohe	Wullaind	Vale				8.485E-07
H_109_40	H_109_40	Consent	Consent	South East	Gohohe	Wullaind '	Male	45	78.9	No	8.485E-07
H_109_44	H_109_44	Consent	Consent	South East	Gohohe	Wullaind '	Male	34	69.2	No	8.485E-07
H_109_51	H_109_51	Absent		South East	Gohohe	Yarraslaic	South				8.485E-07
H_109_55	H_109_55	Consent	Consent	South East	Gohohe	Yarraslaic	Male	32	99.2	Yes	8.485E-07
H_109_55	H_109_55	Absent		South East	Gohohe	Yarraslaic	South				8.485E-07
H_109_58	H_109_58	Absent		South East	Gohohe	Yarraslaic	South				8.485E-07
H_109_59	H_109_59	Consent	Refusal	South East	Gohohe	Yarraslaic	South				8.485E-07

Column 1 (**IID**) contains the **individual unique identifier**, column 2 (**HID**) the **household unique identifier**, column 3 (**HR**) shows **consent/refusal at the household level**, and column 4 (**IR**) **consent/refusal at the individual level**. Columns 5 include the geographic indicators added by default and the variables corresponding to the questions/measurements in the data collection tool, where '**null'** indicates **missing data**.

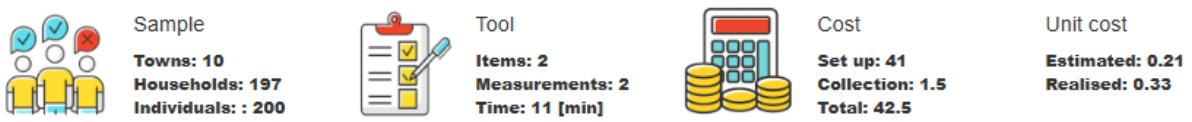
Column 6 (**Probs**) includes the sampling probabilities, calculated by the system based of the sampling strategy.

The second sheet in the downloaded file ("Codebook") provides a list of variable names with their respective content and scale of measurement (categorical/numerical).

1.3. Analysis

The last menu item (**Analysis**), which is only available when survey data have been collected, visualises the results of the data collection process, including:

1) A summary of the sample, data collection tool, and costs (estimated and realised, taking into account response rates)

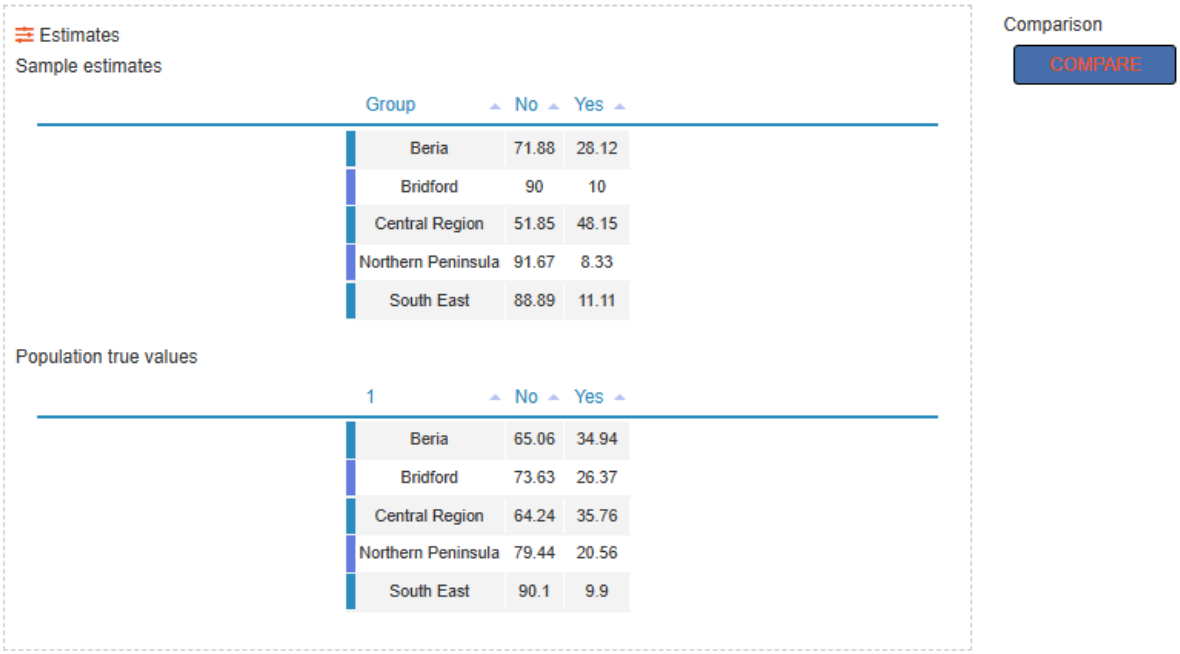


2) A graphical/tabular summary of the response rates (total, by region and by town)

At the bottom, the last section allows for comparison between the raw sample estimates (i.e. simple averages across the sampled individuals) and the 'true' population values, calculated by the system from the full population database.

The comparison is triggered by selecting the outcome of interest, the desired stratification (if any) and the target population, and pressing the COMPARE button:

The results are shown in the **Estimates** panel:



2. Technical Details

With the objective of long-term sustainability, portability to different servers (local or cloud-based) and ease of integration on various learning platforms, the *SurveyLab* is completely developed using open-source, royalty-free technologies.

The basic structure of the system is coded in R v4.5 (R Foundation for Statistical Computing Platform, Vienna, Austria, https://www.R-project.org/) using the Shiny framework (Chang W, Cheng J, Allaire J et al. 2025. Shiny: Web Application Framework for R. R package version 1.11.1.9000, https://github.com/rstudio/shiny). Additional functions for the management of the graphical interface and the visualisation and the dynamic modification of the maps – not directly available in Shiny – are coded in HTML (https://html.spec.whatwg.org/), CSS (https://www.w3.org/Style/CSS/Overview.en.html) and JavaScript https://ecma-international.org/publications-and-standards/standards/ecma-262/).

The Source code of the SurveyLab is available on GitHub:

https://github.com/AnnibaleCois/SurveyLab

It is distributed under MIT license (https://choosealicense.com/licenses/mit/), therefore you are free to use, modify and redistribute the code.

Citing from the licence:

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY,

FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

3. Credits

Annibale Cois Annibale.Cois@mrc.ac.za) conceptualised the SurveyLab, wrote the code of the system and generated the virtual population.

The initial development has been funded through the Fund for Innovation and Research into Teaching and Learning (FIRTL) of Stellenbosch University (FIRLT Grant 30/08/2021). The Research and Social Impact Systems unit at Stellenbosch University supported the engineering of the original idea and coordinated the interaction with the University IT department. The current version has been completely re-written as a ShinyApp and does not include the code developed by the Stellenbosch University IT department.

Acknowledgements:

- The current code includes various JavaScript libraries, which are in the public domain and are distributed with permissive open-source licenses: JQuery (https://jquery.com/), svg-inject (https://github.com/iconfu/svg-inject), panzooom (https://github.com/timmywil/ panzoom), chart.js (https://www.chartjs.org/), tabulator (https://tabulator.info/), g-spinner (https://www.jqueryscript.net/loading/Animated-Circles-Loading-Spinner-jQuery-g-spinner.html).
- The user management relies on a sqlite (https://sqlite.org/) database and – when the *SurveyLab* is deployed on a remote server – on a cloud-based serverless implementation (TursoDB: https://turso.tech/)
- The *SurveyLab* has benefitted by comments, critiques and suggestions by colleagues, friends and students who have used the systems and highlighted deficiencies (or plain errors…) and suggested modifications and improvements.