

WHO Regional Office for Africa | ESPEN

ESPEN-DP-SOP-01

Implementation units, population and endemicity: harmonisation of the JRF and JRSM data

Expanded Special Project for Elimination of Neglected Tropical Diseases (ESPEN)

Document type	Standard operating procedure — data processing
Reference	ESPEN-DP-SOP-01
Version and status	Draft v0.1, for internal review
Date	19 October 2026
Prepared by	ESPEN Data Team, WHO Regional Office for Africa
Scope	List of implementation units (IUs) by year; IU population estimates; harmonisation of the population, endemicity and population requiring PC reported in the JRF and JRSM; endemicity tables by disease, including loiasis
Data sources	ESPEN annual IU boundaries; JRF and JRSM (country information and introduction sheets); UN WPP 2024; WorldPop R2024B; ESPEN endemicity codebook; IU-level loiasis endemicity
Outputs	IU list by year; IU population estimates 2013–2030; IU, country and regional demographic tables 2013–2025; JRF and JRSM submission status by country and year; IU endemicity tables by disease
Processing scripts	Steps 1.1, 3.1, 1.2 and 1.3 (Annex A)
Related documents	SOP-00 (general framework); SOP-02 to SOP-05 (use the outputs of this procedure)

Abbreviations

IU	Implementation unit	RAPLOA	Rapid assessment procedure for loiasis
JAP	Joint Application Package	SAC	School-age children
JRF	Joint Reporting Form	SCH	Schistosomiasis
JRSM	Joint Request for Selected PC Medicines	STH	Soil-transmitted helminthiasis
LF	Lymphatic filariasis	WPP	World Population Prospects (United Nations)
PC	Preventive chemotherapy	WRA	Women of reproductive age
Pre-SAC	Pre-school-age children		

1. Purpose and scope

This procedure describes how ESPEN builds the reference framework of the processing: the list of implementation units (IUs) of each year, the IU population estimates used to complete the populations not reported, and the harmonised IU dataset of population, endemicity and population requiring preventive chemotherapy (PC) reported by countries in the Joint Reporting Form (JRF) and the Joint Request for Selected PC Medicines (JRSM) of the Joint Application Package [1]. It also describes the

split of this dataset into one endemicity table per disease. The outputs are used by all disease-specific procedures.

2. Data sources

Table 1. Data sources

Source	Content used	Years
ESPEN annual IU boundaries	Attributes of each IU (country, administrative levels, IU name, IU code)	2013–2025
JRF and JRSM, country information sheet	IU identifiers, total population and population by age group, endemicity code by disease, population requiring PC by disease and age group, rounds	JRF 2013–2025; JRSM 2015–2025
JRF introduction sheet	National percentages of the population by age group; version of the form	As submitted
UN World Population Prospects 2024 [2]	Annual population growth multiplier and age structure by country	2013–2030
WorldPop gridded population [3]	Population counts at 1 km resolution adjusted to UN totals (release R2024B for 2015–2030; counts prepared for 2013 and 2014)	2013–2030
ESPEN endemicity codebook	Correspondence between the endemicity codes of the forms and standard categories, by disease	Current
IU-level loiasis endemicity	Loiasis endemicity by IU and year, including rapid assessment (RAPLOA) results	As compiled by ESPEN

3. Definitions

Table 2. Terms used in this procedure

Term	Definition
IU-year	One IU in one year of the ESPEN cartography. The IU code and the year form the key of the tables of this procedure.
Population groups	Total population; pre-school-age children (pre-SAC); school-age children (SAC); adults; women of reproductive age (WRA, 15–49 years); children aged 2–4 years. Age groups are those of the JAP forms and of WPP 2024.
Population requiring PC	Population reported by the country as requiring PC for a disease, by age group where the form requests it.
Endemicity category	Standard category of the ESPEN codebook corresponding to the code reported for a disease.
Not reported	IU-year for which neither the JRF nor the JRSM holds a record.
Growth multiplier	Ratio of the national population of a year to that of the previous year (WPP 2024).
Age-group fraction	Share of the national population in an age group, from the JRF introduction sheet or, when not reported, from WPP 2024.

4. Processing steps

4.1 List of implementation units by year

The attribute tables of the annual IU boundary layers are combined into one list, with one record per IU feature and year; the uniqueness of the IU-years is not checked at this step and depends on the boundary files. Leading and trailing spaces are removed, accented characters are converted to plain characters, the names of the administrative levels are written in proper case and numeric fields are stored as whole numbers. No IU is added or removed at this step; the list is the reference to which all reported data are linked.

4.2 IU population estimates

IU population estimates are prepared for each year from 2013 to 2030 and are used only to complete the populations that countries do not report (Section 4.3). They are built as follows:

1. **National parameters.** For each country and year, the growth multiplier is taken from WPP 2024. Pre-SAC, SAC and adult fractions are taken from the JRF introduction sheet; when they are missing in a submitted sheet, they are taken from the previous submitted sheet of the country or, failing that, from the next one. The WRA fraction is not carried over. For any country and year without an introduction sheet (including 2026–2030), and for missing WRA fractions, WPP 2024 fractions are used. The fraction of children aged 2–4 years always comes from WPP 2024.
2. **Reported populations.** The total population reported in the JRF as submitted (before the harmonisation of Section 4.3) is used for an IU and year when it is at least 100; otherwise the total and the age-group values of that record are treated as missing.
3. **Years without a report.** A missing total is estimated from the nearest later reported total divided by the growth multiplier of the year being estimated or, when there is no later value, from the nearest earlier reported total multiplied by that multiplier. The multiplier of a single year is applied, whatever the number of years between the reported and the estimated value.
4. **Age groups.** Reported age-group populations are kept, including values of zero or below 100 when the reported total is valid; missing ones, and the WRA and 2–4-year groups, are obtained by applying the national age-group fraction to the IU total.
5. **Gridded population.** For IUs with no reported population in any year, and for all IUs from 2026 to 2030, the total is the sum of the WorldPop 1 km counts within the IU boundary (each cell weighted by the share of its area inside the IU); the boundaries of 2025 are used for 2026–2030. Age groups are derived with the national fractions. IU-years still without a total after this step are listed for review and left missing.

4.3 Harmonisation of the JRF and JRSM data

The country information sheets of the JRSM and of the JRF are processed in the same way:

1. **Standardisation.** Fields are given standard names, values stored as text are converted to numbers, and names are converted to proper case and plain characters.
2. **Country codes.** The ISO3 code reported for each country is checked against the cartography using four pieces of evidence: the ISO3 code itself, the ISO2 code, the country of the IU codes reported and the country name; a piece of evidence that is missing does not count against the code. A valid reported code is kept, and logged for review when another piece of evidence points to a different country. An invalid code is replaced when all available evidence points to the same country, and left unresolved otherwise; the country name is then left empty, but the records are still linked by IU code. The country name is then replaced by the name used in the cartography.
3. **Endemicity codes.** Known invalid codes are corrected first (for schistosomiasis, code 32 is read as 31 and "ND" as 4, unknown). A missing code is treated as code 0 of the codebook. Codes are

then converted into standard categories with the codebook; codes not found in the codebook are classified as non-endemic.

4. **Missing identifiers.** A JRF record without IU code or IU name takes them from the JRSM of the same country, year and first- and second-level administrative units, only when that unit holds a single IU. This step is applied before the next one, so the completed records take part in it.
5. **Records sharing an IU code.** When several records of the same form have the same IU code and year, records that also have the same administrative units and IU name (ignoring case, accents, spaces and punctuation) are treated as repeats whatever their other values: the record reporting a total population above zero is kept, otherwise the first listed. The remaining records are compared with the cartography entry of that IU code: each is scored from 0 to 1 by the similarity of its administrative units and of its IU name, and its advantage is that score minus its best score with any other IU of the country and year. The record with the largest advantage keeps the IU code if its score is at least 0.5 and its advantage exceeds that of the next record by at least 0.1; otherwise the first record listed (for the JRF, in alphabetical order of administrative units and IU name) keeps it. The other records are kept without IU code, so they are not linked to any IU; the closest other IU is suggested in the log when its score is at least 0.6.
6. **Combination of forms.** Each IU-year of the cartography is linked to the JRF and to the JRSM by IU code and year. Population, endemicity and population requiring PC are taken from the JRF record of the IU-year. Because a blank JRF endemicity code is read as code 0 of the codebook, the JRSM category is used only when the IU-year has no JRF record; a population requiring PC missing in the JRF is taken from the JRSM. Population fields missing or equal to zero in one form are first completed from the other form. IU-years without any record are classified as "Not reported" for LF, onchocerciasis, STH and SCH. Records whose IU code is not in the cartography of the year are not linked.
7. **Completion of populations.** Each population field (total, pre-SAC, SAC, adults) still missing or below 100 is replaced by the IU population estimate of Section 4.2; only the first estimate of each IU and year is used. Because the estimates keep the age-group values reported with a valid total, an age-group value below 100 can remain after this step. National age-group fractions and growth multipliers are attached to each IU-year.
8. **Aggregation.** Populations and populations requiring PC are summed by country and year and by subregion and year; national age-group fractions are carried at country level.
9. **Submission status.** For each country and year, the JRF (or JRSM) is recorded as submitted when the form holds at least one record for the country and year, whether or not it is linked to an IU. Countries that never submitted the form do not appear in the table.

4.4 Endemicity tables by disease

The harmonised dataset is divided into one table per disease (LF, onchocerciasis, loiasis, STH and SCH), each with the IU population, the endemicity category and the population requiring PC of that disease (none for loiasis). All IUs of the countries listed in Table 3 are classified as non-endemic for the disease in every year, irrespective of the category reported.

Table 3. Countries whose IUs are classified as non-endemic in the processing

Disease	Countries
LF	Algeria, Botswana, Burundi, Cabo Verde, Djibouti, Eswatini, Gambia, Lesotho, Mauritania, Mauritius, Namibia, Rwanda, Seychelles, Somalia, South Africa
Onchocerciasis	Algeria, Botswana, Cabo Verde, Comoros, Djibouti, Egypt, Eritrea, Eswatini, Gambia, Lesotho, Madagascar, Mauritania, Mauritius, Namibia, Rwanda, Sao Tome and Principe, Seychelles, Somalia, South Africa, Tanzania (Zanzibar), Zimbabwe

Disease	Countries
STH	Eritrea
SCH	Algeria, Cabo Verde, Comoros, Djibouti, Lesotho, Mauritius, Seychelles

For loiasis, the endemicity of each IU and year is taken from the IU-level loiasis dataset compiled by ESPEN for the same IU and year, which includes RAPLOA results ("RAPLOA 0" is read as non-endemic and "No data" as unknown). When the IU-year is not in that dataset, the category reported by the country is used; IU-years without any information are classified as non-endemic. The highest values of the lower and of the upper RAPLOA bounds recorded for each IU are attached to all its years, and set to zero in the years in which the IU is classified as non-endemic.

5. Outputs

Table 4. Outputs of this procedure

Output	Level	Content
IU list by year	IU, 2013–2025	Identifiers and administrative units of each IU and year
IU population estimates	IU, 2013–2030	Total and age-group populations, growth multipliers and age-group fractions
Demographic tables	IU, country, subregion; 2013–2025	IU: population by age group, age-group fractions and growth multiplier, endemicity category as reported (before the disease-specific rules), population requiring PC by disease and age group, rounds reported in the JRF. Country and subregion: number of IUs, population by age group and population requiring PC by disease and age group (and mean age-group fractions at country level).
Submission status	Country and year	Whether the JRF and the JRSM were submitted
Endemicity tables	IU, 2013–2025	One table per disease, used by SOP-02 to SOP-05 and SOP-07
Decision logs	Record	Country-code inconsistencies; records sharing an IU code and decision taken; IU names that differ from the cartography

6. Quality assurance

- The processing stops if a join between the forms, with the codebook or with the country table would duplicate records, or if records sharing an IU code and year remain after resolution. In the joins with the cartography, the population estimates and the loiasis dataset, only the first record of each IU and year is kept.
- Country-code inconsistencies, records sharing an IU code (with the score and the action taken) and IU names whose similarity with the cartography is below 0.5 are written to log files, which are reviewed after each run; corrections are requested at source when needed.
- Endemicity codes corrected and codes not found in the codebook are reported at each run.

- IU-years without population after completion are cross-tabulated with the endemicity category of each disease.
- The tables of this procedure are written to the analytical database without the empty-table and record-count checks applied to the disease tables (SOP-00, Section 7).
- The list of non-endemic countries (Table 3) and the table of code corrections are reviewed before each reporting cycle.

7. Limitations

- Population estimates are used where countries do not report; they rely on national growth rates and age structures and on gridded population data, and may differ from local census figures.
- Each population field is completed separately, so the age groups of an IU-year may not add up to its total population.
- Codes not found in the codebook are classified as non-endemic; such cases are reported at each run so that the codebook can be completed.
- The country rules of Table 3 override the categories reported by those countries for all years.
- For loiasis, the absence of information is treated as non-endemic.
- Records that cannot be attributed to an IU are not included in the tables until the IU code is corrected at source. Records whose IU code does not exist in the cartography of the year are not linked and are not currently listed in the decision logs.
- The IU tables do not show which populations were estimated: the field marking IU-years without population is calculated after completion.
- The population estimates use the JRF records as submitted, before the resolution of records sharing an IU code; an IU-year can therefore have more than one estimate, of which the first is used.

8. Revision

This procedure describes the processing as implemented in the versions of the R scripts listed in Annex A (October 2026). It is revised whenever a decision rule, threshold, category or data source changes, and at least once a year before the new reporting cycle. Each revision is recorded in the table below.

Version	Date	Changes
0.1	19 October 2026	First draft of SOP-01, prepared from the code of the processing scripts.

References

1. World Health Organization. Joint application package for preventive chemotherapy [Internet]. Geneva: World Health Organization; [cited 2026 Oct 19]. Available from: <https://www.who.int/teams/control-of-neglected-tropical-diseases/interventions/strategies/preventive-chemotherapy/joint-application-package>
2. United Nations, Department of Economic and Social Affairs, Population Division. World Population Prospects 2024 [Internet]. New York: United Nations; 2024 [cited 2026 Oct 19]. Available from: <https://population.un.org/wpp/>
3. WorldPop. Global population estimates 2015–2030, release R2024B, 1 km resolution, UN-adjusted [Internet]. Southampton: University of Southampton; [cited 2026 Oct 19]. Available from: <https://www.worldpop.org>

Annex A. Processing scripts

Step	Script	Role in this procedure
1.1	Processing_IU_list_v02.R	List of IUs by year (Section 4.1)
3.1	Population_estimates_v03.R	IU population estimates (Section 4.2); run before step 1.2 whenever the JRF data, the boundaries or the gridded population change
1.2	ESPEN_Indicators_Demographics_v04.R	Harmonisation of the JRF and JRSM data (Section 4.3)
1.3	ESPEN_Indicators_Endemicity_v02.R	Endemicity tables by disease (Section 4.4)
–	Functions/JAP_functions.R	Shared functions: country-code validation, endemicity recoding, population completion, resolution of records sharing an IU code