

WHO Regional Office for Africa | ESPEN

ESPEN-DP-SOP-00

General framework for processing Joint Application Package data into the ESPEN analytical tables

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

Document type	Standard operating procedure — data processing
Reference	ESPEN-DP-SOP-00
Version and status	Draft v0.1, for internal review
Date	19 October 2026
Prepared by	ESPEN Data Team, WHO Regional Office for Africa
Scope	All data processed by ESPEN to produce the analytical tables of the ESPEN NTD Portal for lymphatic filariasis, onchocerciasis, loiasis, soil-transmitted helminthiases, schistosomiasis and trachoma
Data sources	Joint Reporting Form (JRF) and Joint Request for Selected PC Medicines (JRSM); GET2020 trachoma database; ESPEN IU boundaries; population, WASH and reference datasets
Outputs	Analytical tables at implementation unit (IU), country and regional level; cartography tables; map and chart specifications
Processing scripts	R scripts of the ESPEN data team, October 2026 versions (Annex A)
Related documents	SOP-01 to SOP-08 (disease-specific and cross-cutting procedures)

Abbreviations

AT	Analytical table (processed IU, country or region dataset)	LF	Lymphatic filariasis
EU	Evaluation unit	MDA	Mass drug administration
GET2020	WHO Alliance for the Global Elimination of Trachoma by 2020 (database held by ITI)	NTD	Neglected tropical disease
IHME	Institute for Health Metrics and Evaluation	PC	Preventive chemotherapy
ITI	International Trachoma Initiative	PC-NTD	NTD amenable to preventive chemotherapy
IU	Implementation unit	SAC	School-age children
JAP	Joint Application Package	SCH	Schistosomiasis
JMP	WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene	STH	Soil-transmitted helminthiases
JRF	Joint Reporting Form	WASH	Water, sanitation and hygiene
JRSM	Joint Request for Selected PC Medicines	WPP	World Population Prospects (United Nations)

1. Purpose and scope

The Expanded Special Project for Elimination of Neglected Tropical Diseases (ESPEN) of the WHO Regional Office for Africa compiles the data reported by national programmes on the neglected tropical diseases amenable to preventive chemotherapy (PC-NTDs): lymphatic filariasis (LF), onchocerciasis, schistosomiasis (SCH), soil-transmitted helminthiasis (STH) and trachoma, together with loiasis, which conditions the choice of treatment for filarial diseases. These data are processed into analytical tables (ATs) that feed the dashboards, maps and country pages of the ESPEN NTD Portal [1] and support the monitoring of the road map for neglected tropical diseases 2021–2030 [2].

This document describes the general framework of the processing: data sources, geographic framework, processing cycle, conventions common to all diseases, quality assurance, publication and interpretation. The disease-specific rules and the cross-cutting components are described in the procedures listed in Table 1. The procedures describe the processing as implemented in the R scripts of the ESPEN data team in October 2026; rules are stated as they are applied by the code.

Table 1. ESPEN data-processing procedures

Code	Title	Content
SOP-00	General framework	Sources, geographic framework, processing cycle, common conventions, quality assurance, publication
SOP-01	Implementation units, population and endemicity	IU list by year, IU population estimates, harmonisation of JRF and JRSM data, endemicity tables by disease, loiasis
SOP-02	Lymphatic filariasis	Endemicity, MDA rounds, coverage and summary indicators
SOP-03	Onchocerciasis	Endemicity, MDA rounds, coverage, elimination-mapping flag and summary indicators
SOP-04	Soil-transmitted helminthiasis	Endemicity, population requiring PC and PC by age group, coverage and summary indicators
SOP-05	Schistosomiasis	Endemicity, population requiring PC by reporting scheme, PC by age group, coverage and summary indicators
SOP-06	Trachoma	Survey results, MDA with azithromycin, coverage and summary indicators from the global trachoma database
SOP-07	Co-endemicity, WASH and country summaries	Access to water and sanitation by IU, co-endemicity and recommended MDA strategies, country and regional summaries across PC-NTDs
SOP-08	Cartography and portal specifications	Cartography tables, IU map layers, maps available in the Map Generator Tool, dashboard charts

2. Overview of the data flow

The processing is organised in successive stages that run in a fixed order, each stage using the outputs of the previous one (Figure 1):

1. **Reference framework.** List of IUs for each year from the official IU boundaries, and IU population estimates used to complete the populations not reported (SOP-01).

2. **Harmonisation of the JAP data.** Population, endemicity and population requiring PC reported in the JRF and the JRSM are standardised and linked to the IU list of each year (SOP-01).
3. **Disease-specific processing.** Endemicity categories, treatment rounds, coverage and summary indicators are produced for each disease at IU, country and regional level (SOP-02 to SOP-06).
4. **Integration.** The disease tables are finalised and combined into co-endemicity tables with recommended MDA strategies, country summaries across PC-NTDs and yearly IU map layers (SOP-07, SOP-08).
5. **Publication.** The ATs are stored in the ESPEN analytical database, from which the portal products are generated (Section 8).

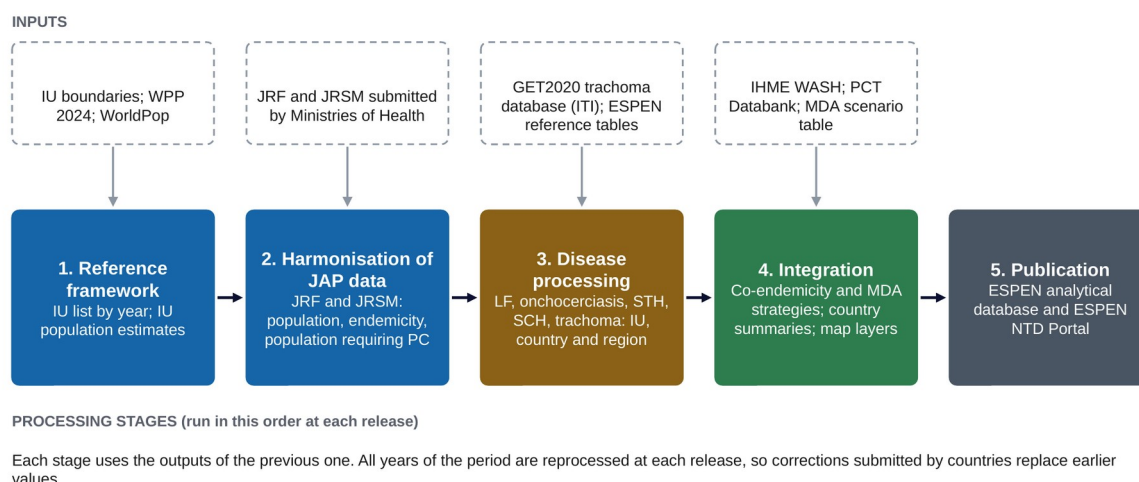


Figure 1. Stages of the processing of JAP and related data into the ESPEN analytical tables.

3. Data sources

Table 2 lists the data sources. The JAP forms are the main source: they are completed by national programmes and submitted by Ministries of Health to WHO [3]. Trachoma data are not reported through the JRF; they are obtained from the global trachoma database maintained by the International Trachoma Initiative (ITI) on behalf of national programmes [4].

Table 2. Data sources

Source	Content used	Provider	Used in
Joint Reporting Form (JRF)	IU population by age group, endemicity status, population requiring PC, PC delivered by regimen and age group (population targeted, population treated, coverage), version of the form	Ministries of Health [3]	SOP-01 to SOP-05, SOP-07
Joint Request for Selected PC Medicines (JRSM)	IU population, endemicity status, population requiring PC, planned rounds	Ministries of Health [3]	SOP-01, SOP-07
Global trachoma database (GET2020)	Trachoma surveys (TF and TT categories and prevalence), azithromycin distribution and coverage, surgery, IU population	ITI [4]	SOP-06
ESPEN IU boundaries	IU, first administrative level and country boundaries for each year since 2013	ESPEN, from boundaries submitted by countries	All

Source	Content used	Provider	Used in
Population data	Annual growth rates and age structure; gridded population counts at 1 km	UN WPP 2024 [5]; WorldPop R2024B [6]	SOP-01
WASH estimates	Household access to water and sanitation, 2000–2017; country indicators	IHME [7]; JMP [8]	SOP-07
PCT Databank	Country-level status of LF, onchocerciasis, STH and SCH	WHO [9]	SOP-07
ESPEN reference tables	Endemicity codebook; contemporary IU endemicity (2020), updated at a regional workshop with national programme and data managers at the end of 2019; historical MDA records (APOC) for LF and onchocerciasis; list of redistricted IUs; IU-level loiasis endemicity; scenarios of recommended MDA strategies; map and chart specifications	ESPEN	As stated in each SOP

4. Geographic framework

The implementation unit (IU) is the unit of analysis of the JAP data. An IU is the administrative unit, usually a district, in which MDA is planned and implemented. Each year the IU boundaries reported by countries are incorporated in the ESPEN cartography, and each IU carries a unique IU code for that year. The JRF and JRSM records are linked to the IU list of the corresponding year through the IU code and the year. Administrative names are also used in two cases: to complete the IU code of JRF records that have none (from the JRSM, when the administrative unit holds a single IU), and to give the treatment records the IU code of the JRF country information sheet with the same country, first- and second-level administrative units and year (SOP-01, SOP-02).

IU boundaries change when units are split or merged (redistricting). ESPEN keeps a list of IUs created by redistricting, with their parent IU and the year of creation. An IU created by a split from 2014 onwards inherits the cumulative counts of its first parent IU as of the year before its creation, on the assumption that treatment was delivered uniformly across the parent IU: the parent's rounds since 2013 are added to its own and, for LF and onchocerciasis, the cumulative total including historical rounds starts from the larger of its own and its parent's value (SOP-02 to SOP-05).

Trachoma data are processed on the IU geography of the global trachoma database, which may differ from the ESPEN IUs; trachoma tables are linked to the trachoma boundaries through the identifier of that database (SOP-06).

Results are aggregated by country and by United Nations subregion within each WHO region (Western, Middle, Eastern, Southern and Northern Africa, and Western Asia for the countries of the WHO Eastern Mediterranean Region covered by the ESPEN cartography). The disease tables do not include a total for the whole region; totals for the WHO African Region and for all countries covered are produced in the country summaries across PC-NTDs (SOP-07).

5. Processing cycle and versioning

The processing is run when new or corrected submissions are to be reflected in the portal. At each run all years of the period covered (currently 2013–2025 for the IU tables) are processed again from the source data. A correction submitted by a country for an earlier year therefore replaces the previous value at the next release, and published figures can change between releases.

Each release is identified by the year and month of the run (YYYYMM), which is used in the names of the exported copies of the analytical tables. The steps pass their results to each other through working files and database tables that are replaced at each run. The processing scripts are versioned; this series of procedures refers to the versions listed in Annex A. Data are organised by IU and year; JRSM records are linked to the JRF records of the same IU and year.

6. Conventions common to all diseases

The conventions in Table 3 apply to all diseases unless the disease-specific procedure states otherwise. Definitions of coverage follow WHO guidance on PC data [10,11]; the thresholds for an effective round correspond to the coverage targets of the WHO programmes [12–14].

Table 3. Conventions applied in the processing

Topic	Rule applied
Hierarchy of sources	For each IU and year the JRF record is used when there is one. Its endemicity category is used even when the code was left blank (it is then read as code 0 of the codebook), so the JRSM category is used only for IU-years without a JRF record. Population fields missing or zero in one form are completed from the other form, and a population requiring PC missing in the JRF is taken from the JRSM. Populations still missing, or below 100, are replaced by IU population estimates (SOP-01).
Linkage	Records are linked to the IU list by IU code and year. Names are used to resolve records that share an IU code, to complete missing JRF IU codes from the JRSM (only for administrative units with a single IU), and to give treatment records the IU code of the country information sheet with the same country, first- and second-level administrative names and year.
Country codes	Reported country codes are checked against the cartography using the ISO3 code, the ISO2 code, the IU codes and the country name. A valid reported ISO3 code is kept, and logged when other evidence contradicts it; an invalid code is replaced when all available evidence agrees. Country names follow the ESPEN cartography, with four names harmonised (Democratic Republic of the Congo, Sao Tome and Principe, Gambia, Cabo Verde).
Endemicity categories	Reported codes are converted into standard categories with the ESPEN endemicity codebook. Blank codes are read as code 0 of the codebook, and codes not found in the codebook are classified as non-endemic. IU-years without any report are classified as "Not reported" for LF, onchocerciasis, STH and SCH, and as non-endemic for loiasis.
Duplicated records	Records of the same form with the same IU code, year, administrative units and IU name (ignoring case, accents, spaces and punctuation) are treated as repeats whatever their other values: the record that reports a total population is kept, otherwise the first listed. Treatment data keep one record per IU, year and regimen (the record with the largest population treated) and one record per IU and year after the regimens are combined; populations are not added across regimens or rounds, and the largest value reported is kept (for onchocerciasis, the ALB+IVM record when reported, otherwise ivermectin alone).
Programme coverage	Population treated divided by population targeted, $\times 100$.
Epidemiological coverage	Population treated divided by population requiring PC, $\times 100$.

Topic	Rule applied
Denominators	When people were treated but the population targeted is missing or very small, the population requiring PC is used as the denominator, and coverage is set to zero when its denominator is below a minimum. For STH and SCH, the population treated is first limited to the population requiring PC of each age group. The thresholds are given in each disease procedure.
Treatment round	An IU-year counts as one round when coverage is above zero. The tables hold one round per IU and year, whatever the number of rounds delivered (for the exceptions in trachoma, see SOP-06).
Effective round	Coverage at or above the programme threshold: 65% for LF, 80% for onchocerciasis, 75% (total programme coverage) for STH and SCH. No threshold is applied for trachoma.
Cumulative rounds	Number of years with a round (or an effective round) for each IU since 2013. For LF and onchocerciasis, a separate cumulative total is the number of rounds in the historical (APOC) records, minus the round of 2014 (which these records are assumed to include), plus the rounds counted since 2013. New IUs inherit the counts of their parent IU (Section 4).
Aggregation	Country and subregional values are sums over the IUs in the relevant categories. Coverage above IU level is calculated from the summed populations, not averaged across IUs.
Rounding	Coverage and RAPLOA values are stored with two decimals, age-group fractions with three and growth multipliers with five. In the disease tables, other values are rounded to whole numbers (to the nearest even number when a value ends in exactly 0.5); in the IU list and the demographic tables, the decimal part is dropped.
Missing values	Missing populations and treatment values are set to zero at several steps. A zero in the tables therefore does not always distinguish a value not reported from a value reported as zero.

7. Quality assurance

Quality assurance is built into the processing and completed by a review of the decision logs after each run:

- **Automatic stops.** The processing stops when a join between the forms, with the codebook or with the country table would duplicate records, when an expected table or field is missing and, for the disease tables and the portal specifications, when an empty table would replace a published one. Joins with the cartography and with the population estimates keep the first record of each IU and year instead of stopping.
- **Decision logs.** Inconsistent country codes, records that share an IU code with the decision taken, and records whose IU name differs from the cartography are written to log files. The ESPEN data team reviews them and, when needed, asks for the correction at source.
- **Record counts.** For the disease tables and the portal specifications, the number of records stored in the analytical database is compared with the number prepared.
- **Consistency checks.** Reported and processed endemicity categories are cross-tabulated, the number of records edited is reported, and categories without a standard code, scenarios without a recommended MDA strategy and records left without an IU code are listed. Records whose IU code is not found in the cartography of the year are not yet listed (SOP-01).

- **Reference tables.** The parameters and reference tables listed in each procedure are reviewed before each reporting cycle.

8. Publication on the ESPEN NTD Portal

The ATs are stored in the ESPEN analytical database: IU, country and regional tables for each disease, co-endemicity tables, country summaries across PC-NTDs, cartography tables and the lookup tables that define the maps and charts. The ESPEN NTD Portal [1] reads these tables to produce the dashboards, the Map Generator Tool, the country pages and the data downloads. The maps available for each country and period are defined as described in SOP-08.

9. Interpretation and limitations

- The tables reflect the data reported by national programmes. The processing harmonises, links and completes the data but does not verify the accuracy of field data.
- Validated country data take precedence over estimates. Where a country does not report a value, estimates are used as described in SOP-01; these are identified as estimates in the procedures but not flagged record by record in the tables.
- Some rules reclassify reported values (for example, countries considered non-endemic for a disease, or the reconciliation of 2013–2019 endemicity with the contemporary classification agreed with national programme and data managers at the end of 2019). The category reported by the country, after conversion with the codebook, is kept in the IU demographic table; the category before reconciliation is kept in the intermediate processing files but not in the published disease tables.
- Programme coverage can exceed 100% when the population treated exceeds the population targeted.
- Figures for the most recent year may be incomplete, because reports are received during the following year.
- ESPEN figures can differ from those published by WHO headquarters, such as the PCT Databank [9], because of differences in timing, in the level of processing (IU rather than country) and in the completion of missing values.

10. Responsibilities and revision

The ESPEN data team runs the processing, reviews the decision logs, maintains the reference tables and keeps this series of procedures up to date. National programmes are contacted when data need to be corrected at source.

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-00, prepared from the code of the processing scripts.

References

1. World Health Organization Regional Office for Africa. ESPEN NTD Portal [Internet]. Brazzaville: WHO Regional Office for Africa; [cited 2026 Oct 19]. Available from: <https://espen.afro.who.int>

2. World Health Organization. Ending the neglect to attain the Sustainable Development Goals: a road map for neglected tropical diseases 2021–2030. Geneva: World Health Organization; 2020. Available from: <https://www.who.int/publications/i/item/9789240010352>
3. 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>
4. International Trachoma Initiative. Trachoma data and the global trachoma database [Internet]. Decatur (GA): Task Force for Global Health; [cited 2026 Oct 19]. Available from: <https://www.trachoma.org>
5. 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/>
6. 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>
7. Local Burden of Disease WaSH Collaborators. Mapping geographical inequalities in access to drinking water and sanitation facilities in low-income and middle-income countries, 2000–17. *Lancet Glob Health*. 2020;8(9):e1162–85. doi:10.1016/S2214-109X(20)30278-3
8. WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene. JMP WASH data [Internet]. [cited 2026 Oct 19]. Available from: <https://washdata.org>
9. World Health Organization. PCT databank [Internet]. Geneva: World Health Organization; [cited 2026 Oct 19]. Available from: <https://www.who.int/teams/control-of-neglected-tropical-diseases/data-platforms/pct-databank>
10. World Health Organization. Preventive chemotherapy: tools for improving the quality of reported data and information. A field manual for implementation. Geneva: World Health Organization; 2019. Available from: <https://www.who.int/publications/i/item/9789241516464>
11. World Health Organization. Preventive chemotherapy in human helminthiasis: coordinated use of anthelmintic drugs in control interventions. A manual for health professionals and programme managers. Geneva: World Health Organization; 2006. Available from: <https://www.who.int/publications/i/item/9241547103>
12. World Health Organization. Monitoring and epidemiological assessment of mass drug administration in the global programme to eliminate lymphatic filariasis: a manual for national elimination programmes. Geneva: World Health Organization; 2011. Available from: <https://www.who.int/publications/i/item/9789241501484>
13. World Health Organization. Guidelines for stopping mass drug administration and verifying elimination of human onchocerciasis: criteria and procedures. Geneva: World Health Organization; 2016. Available from: <https://www.who.int/publications/i/item/9789241510011>
14. World Health Organization. Guideline: preventive chemotherapy to control soil-transmitted helminth infections in at-risk population groups. Geneva: World Health Organization; 2017. Available from: <https://www.who.int/publications/i/item/9789241550116>

Annex A. Processing scripts documented

The procedures document the following R scripts (October 2026 versions). Scripts are listed in order of execution; the step numbers are those of the ESPEN workflow.

Step	Script	Folder	SOP	Stage
1.1	Processing_IU_list_v02.R	Processing_IU_list	SOP-01	Reference framework
3.1	Population_estimates_v03.R	Processing_Population	SOP-01	Reference framework
1.2	ESPEN_Indicators_Demographics_v04.R	Processing_Population	SOP-01	Harmonisation
1.3	ESPEN_Indicators_Endemicity_v02.R	Processing_Endemicity	SOP-01	Harmonisation
1.4	LF_AT_v05.R	Processing_LF	SOP-02	Disease processing
1.5	Oncho_AT_v05.R	Processing_Oncho	SOP-03	Disease processing
1.6	STH_AT_v05.R	Processing_STH	SOP-04	Disease processing
1.7	SCH_AT_v05.R	Processing_SCH	SOP-05	Disease processing
1.8	Trachoma_AT_v05.R	Processing_Trachoma	SOP-06	Disease processing
1.9	ESPEN_AT_Transpose_v02.R	Processing_AT	SOP-02 to SOP-06	Integration
2.1	Processing_WASH.R	Processing_WaSH	SOP-07	Integration
1.10	Coendemicity_AT_v02.R	Processing_Coendemicity	SOP-07	Integration
1.11	Country_Coendemicity_AT_v03.R	Processing_Coendemicity	SOP-07	Integration
1.12	IU_AT_shp_v02.R	Processing_IU_SHP	SOP-08	Integration
5.1	IU_AT_GeoJSON.R	Processing_IU_SHP	SOP-08	Publication
6.1	Maps_Specs_Categories_v02.R	Processing_Map_Chart_specs	SOP-08	Publication
6.2	Maps_Specs_v04.R	Processing_Map_Chart_specs	SOP-08	Publication
6.3	Chart_Specs_v02.R	Processing_Map_Chart_specs	SOP-08	Publication
–	Functions/JAP_functions.R	Functions	All	Shared functions