

ScanForm Form Documentation

Form Documentation

Table of contents

| | | |
|-----------|---|-----------|
| 1 | TB Facility Register v.5 | 3 |
| 2 | About this form | 4 |
| 3 | What data does this form collect? | 5 |
| 4 | How ScanForm digitises this form | 6 |
| 5 | Data quality: a three-layer system | 7 |
| 6 | Documentation tabs | 8 |
| 7 | Correct completion instructions | 9 |
| 8 | About ScanForm | 10 |
| 9 | Data Pipeline | 11 |
| 10 | Pipeline Overview | 12 |
| 11 | Layer-by-Layer Description | 13 |
| 11.1 | Base - mw_tb_base_mat_pip_MW_TB_Register_v1 | 13 |
| 11.2 | Pre-Clean - Two models | 13 |
| 11.2.1 | mw_tb_pre_clean__tb_reg_raw | 13 |
| 11.2.2 | mw_tb_pre_clean__tb_reg | 13 |
| 11.3 | Clean - Two models | 15 |
| 11.3.1 | mw_tb_clean__tb_reg_checks | 15 |
| 11.3.2 | mw_tb_clean__tb_reg | 15 |
| 11.4 | Intermediary - Two models | 16 |
| 11.4.1 | mw_tb_intermediary__tb_aggregated | 16 |
| 11.4.2 | mw_tb_intermediary__checks_with_locations | 16 |
| 11.5 | Refinery - Two models | 16 |
| 11.5.1 | mw_tb_refinery__tb_reg | 16 |
| 11.5.2 | mw_tb_refinery__tb_reg_checks | 17 |
| 12 | Key Design Decisions | 18 |
| 13 | Model Inventory | 19 |
| 14 | Variables | 20 |
| 15 | Overview | 21 |
| 16 | Field Types at a Glance | 22 |

| | |
|--|-----------|
| 17 Variables by Section | 23 |
| 18 Full Variable Reference | 24 |
| 18.1 Patient Identity | 24 |
| 18.2 Registration | 24 |
| 18.3 Demographics | 24 |
| 18.4 Risk Group | 25 |
| 18.5 Referral | 25 |
| 18.6 Treatment | 25 |
| 18.7 Initial Lab Results | 26 |
| 18.8 Follow-up Smears | 26 |
| 18.9 Imaging and DST | 27 |
| 18.10 Treatment Outcome | 27 |
| 18.11 HIV and ART | 28 |
| 18.12 Post-TB and PRP | 28 |
| 18.13 Other | 29 |
| 18.14 Administrative | 29 |
| 19 Conditional Logic Map | 30 |
| 20 OCR Models Used | 31 |
| 21 Required vs Optional Fields | 32 |
| 22 In-Phone Validation | 33 |
| 23 How in-phone validation works | 34 |
| 24 Core concepts | 35 |
| 25 Eligibility rules | 36 |
| 25.1 Page-level gate | 36 |
| 25.2 Row-level gate | 36 |
| 26 Validated fields and checks | 38 |
| 27 Check types explained | 39 |
| 28 Fields not validated in-phone | 40 |
| 29 Validation summary | 41 |
| 30 Data Quality Checks | 42 |
| 31 Overview | 43 |
| 32 Severity levels | 44 |
| 33 Check catalogue | 45 |
| 33.1 Integer conversion checks | 45 |
| 33.2 Date conversion checks - Warnings | 45 |
| 33.3 Date conversion checks - Errors | 46 |
| 33.4 Multiple-answer selection checks | 46 |

| | |
|--|-----------|
| 33.5 Clinical logic check | 47 |
| 34 All checks at a glance | 49 |
| 35 Checks output table structure | 50 |
| 36 Relationship to in-phone validation | 51 |
| 37 Dashboard Indicators | 52 |
| 38 Overview | 53 |
| 39 Indicator catalogue | 54 |
| 40 Programme Performance | 55 |
| 40.1 Total patients registered | 55 |
| 40.2 Treatment success rate | 55 |
| 40.3 Treatment outcome distribution | 55 |
| 40.4 Lost to follow-up rate | 56 |
| 40.5 Mortality rate during treatment | 56 |
| 40.6 Moved to second-line treatment | 56 |
| 40.7 Patient age and sex profile | 57 |
| 40.8 Occupation and risk group distribution | 57 |
| 40.9 Previously treated patients | 57 |
| 41 Bacteriology & Diagnosis | 59 |
| 41.1 Bacteriological confirmation rate | 59 |
| 41.2 Xpert MTB positivity rate | 59 |
| 41.3 Drug-resistant TB rate (DST) | 59 |
| 41.4 Smear conversion rate at 2 months | 60 |
| 42 Treatment Regimen | 61 |
| 42.1 Treatment regimen distribution | 61 |
| 42.2 DOT and adherence support coverage | 61 |
| 43 HIV & ART | 62 |
| 43.1 HIV co-infection rate among TB patients | 62 |
| 43.2 HIV testing coverage among TB patients | 62 |
| 43.3 ART initiation among HIV-positive TB patients | 62 |
| 43.4 CPT coverage among HIV-positive TB patients | 63 |
| 44 Post-TB / PRP | 64 |
| 44.1 PRP enrolment rate among eligible patients | 64 |
| 44.2 PTLD diagnosis method distribution | 64 |
| 44.3 PRP programme outcome distribution | 64 |
| 45 Data Quality | 66 |
| 45.1 Check failure rate by book | 66 |
| 45.2 Most frequent check failures | 66 |
| 45.3 Records excluded due to Errors | 66 |
| 46 Indicator summary table | 68 |

1 TB Facility Register v.5

Form Documentation

i Note

This is live documentation. It is automatically regenerated each time the underlying XLSForm, dbt pipeline, or validation configuration changes. The content on every tab always reflects the currently deployed form.

2 About this form

The **TB Facility Register (v.5)** is a paper-based data collection instrument used at health facilities in **Malawi** to record the registration, diagnosis, treatment, and outcomes of tuberculosis patients. It is issued by the **Ministry of Health** and digitised using **ScanForm** — an optical character recognition (OCR) platform developed by QED that converts photographed paper forms into structured electronic data.

Each physical form is printed on an **840 × 297 mm** wide-format sheet (a double-A4 landscape spread) and accommodates **5 patient records per page**, arranged in rows. The left page captures patient identity, registration, diagnosis, and initial laboratory results; the right page captures follow-up smear results, drug susceptibility testing, treatment outcomes, HIV and ART details, and post-TB lung disease programme data.

Table 2.1: Form at a glance

| Attribute | Value |
|---------------------------|------------------------------------|
| Programme | MW-TB |
| Version | v.5 |
| XLSForm release | 2026-03-02-v1.0.0 |
| Country | Malawi |
| Issuing body | Ministry of Health |
| Paper size | 840 x 297 mm (double-A4 landscape) |
| Patients / page | 5 rows |
| Total variables | 56 per patient row |
| Mandatory fields | 11 |
| Conditional fields | 9 |
| Digitisation | ScanForm OCR (QED.ai) |

3 What data does this form collect?

The register follows a TB patient from first contact through to final outcome. The sections below summarise the thematic groups of data collected.

Table 3.1: Data sections on the TB Facility Register (61 variables across 12 sections)

| Section | Fields | Key Variables |
|----------------------------------|--------|--|
| Patient Identity | 3 | Full name, address, phone number |
| Registration | 3 | Registration date, Participant ID, TB registration number |
| Demographics | 3 | Age (years or months), sex |
| Risk Group & Referral | 4 | Occupation / risk group (8 categories including miners), referral source |
| Treatment | 5 | Regimen (2RHZE/4RH, BPalM, BPal), site of disease, treatment history, diagnosis method |
| Initial Lab Results | 10 | LAM, Smear, Xpert, Culture - dates and MTB detected/not detected results |
| Follow-up Smears | 9 | 2-month, 5-month, 6-month smear dates and graded results |
| Imaging & DST | 4 | X-ray result, DST result (RIF resistance), dates |
| Treatment Outcome | 4 | Outcome (6 categories), outcome date, 2nd line escalation, DOT support |
| HIV & ART | 7 | HIV status, test timing, ARV start time, ART ID, CPT status and start date |
| Post-TB / PRP | 3 | PRP enrolment, PTLD diagnosis method, PRP outcome |
| Comments & Admin | 6 | Free-text comments, book/page numbers, discard and error flags |

4 How ScanForm digitises this form

ScanForm uses **Optical Character Recognition (OCR)** to convert photographed paper forms into structured data. Each field on the paper is assigned an OCR model tuned to the type of input expected.

Table 4.1: OCR input types on this form

| Input Type | How OCR Processes It | Example Fields |
|--------------------------|--|---|
| Digit boxes | Each box is recognised independently as a digit 0-9 using the int model. | Age, smear result, registration date |
| Letter boxes | Each box is recognised as A-Z. Restricted alphabets reduce errors for known value sets. | Age unit (Y/M), ID codes |
| Oval bubbles | The select one or zero model detects whether each oval has an X crossing. At most one positive per question. | Sex, HIV status, treatment outcome |
| Pre-printed text | Some boxes contain fixed printed characters (e.g. MTB-, ART-, 20). These are not OCR-scanned. | TB registration number prefix, year century |
| Handwriting zones | Free-text lines are not OCR-processed. They are captured as images for human review. | Full name, address, comments |

5 Data quality: a three-layer system

Data quality is enforced at three distinct stages, each acting as a safety net for the next.

Table 5.1: Three-layer data quality system

| Stage | Layer | When | Who Acts | What Happens |
|-------|-----------------------------|--|------------------|---|
| 1 | In-phone validation | At photo capture, before submission | Field worker | 26 key fields checked per row. Missing required fields and multiple ovals trigger an immediate on-screen alert. Worker corrects the paper form and retakes the photo. |
| 2 | Pipeline DQA checks | After submission, on each pipeline run | Data manager | 41 fields checked across 5 check types: integer parse, date parse, multiple-answer selection, and clinical logic. Errors exclude records; Warnings flag for review. |
| 3 | Dashboard monitoring | Ongoing, daily or monthly | Supervisor / MoH | 25 programme indicators surfaced in dashboards. Trends, outliers and facility comparisons make systemic data quality issues visible at programme level. |

6 Documentation tabs

The online documentation site organises content into the tabs described below. This PDF consolidates the same material into chapters.

Table 6.1: Documentation site map

| Tab | Primary Audience | Contents |
|-----------------------------|-----------------------------------|---|
| Home | <i>All</i> | Form summary, data collection overview, and orientation guide. |
| Variables | <i>Data managers, researchers</i> | Complete field-by-field reference: variable names, input types, required status, conditional logic, OCR models, and constraints for all 56 variables. |
| Data Pipeline | <i>Technical staff, analysts</i> | How raw OCR output flows through 5 dbt transformation layers (Base to Pre-Clean to Clean to Intermediary to Refinery) to produce analysis-ready tables. |
| In-Phone Validation | <i>Field supervisors</i> | The 26 automated checks that run instantly on the field worker's phone when a page is photographed. Errors prompt immediate on-site correction. |
| DQA Checks | <i>Data managers</i> | The 5 server-side data quality check blocks that run after submission: integer validation, date validation, multiple-selection detection, and clinical logic. |
| Dashboard Indicators | <i>Programme managers, MoH</i> | 25 actionable indicators derivable from this form, with numerators, denominators, disaggregations, and recommended visualisations. |

7 Correct completion instructions

How to fill in this form

These instructions are printed on the back of the physical form.

Filling out

- Use the square boxes for **numbers and letters**.
- Write exactly **one letter or digit per box**.
- Use **CAPITAL letters** only.
- Cross an oval bubble with an **X** to select an option.
- Questions marked with an **asterisk** are mandatory.

Correcting mistakes

- Shade the incorrect answer completely.
- Write the correct answer **close to it**, outside the original box.
- Do not use correction fluid.

Example: If *W* was written incorrectly, shade it and write the correct letter nearby.

Treatment History codes

| Code | Full Meaning |
|----------------|---|
| New | New case - never previously treated |
| Relapse | Previously treated - cured or completed treatment |
| RALF | Return after lost to follow-up |
| Fail | Treatment after treatment failure |
| Other | Previous treatment outcome unknown |
| Unknown | Treatment history unknown |

8 About ScanForm

[ScanForm](#) is a paper-based data digitisation platform developed by **QED.ai** for use in international development and public health programmes. Forms are designed in an extended XLSForm standard (XLS-ScanForm), printed and used in the field in the usual way, then photographed with a smartphone. The ScanForm mobile app applies OCR models to recognise handwritten digits, letters, and marked ovals, and immediately validates the captured data before uploading to the cloud.

i Note

Document version: Auto-generated from XLSForm MW-TB_TB_Register.xlsx (release 2026-03-02-v1.0.0). For questions about this documentation contact the QED data team.

9 Data Pipeline

10 Pipeline Overview

Scanned forms are ingested from ScanForm, exported as a materialised dataset, and then processed through five transformation layers before reaching dashboards and reporting tools. Each layer has a specific responsibility: ingestion, filtering, cleaning, aggregation, and final delivery.

Table 10.1: Pipeline data flow across 5 layers

| Step | Layer | Model | Reads From |
|------|--------------|--|------------------------------|
| 1 | Source | mat pip MW TB Register v1 | ScanForm materialised export |
| 2 | Base | mw tb base mat pip MW TB Register v1 | Source |
| 3a | Pre-Clean | mw tb pre clean tb reg raw | Base |
| 3b | Pre-Clean | mw tb pre clean tb reg | Pre-Clean (raw) |
| 4a | Clean | mw tb clean tb reg checks | Pre-Clean (raw) |
| 4b | Clean | mw tb clean tb reg | Pre-Clean + checks |
| 5a | Intermediary | mw tb intermediary tb aggregated | Clean tb reg |
| 5b | Intermediary | mw tb intermediary checks with locations | Clean checks + locations |
| 6a | Refinery | mw tb refinery tb reg | Clean tb reg |
| 6b | Refinery | mw tb refinery tb reg checks | Clean tb reg checks |

11 Layer-by-Layer Description

11.1 Base - `mw_tb_base__mat_pip_MW_TB_Register_v1`

The base model reads directly from the **ScanForm materialised export** (`mat_pip_MW_TB_Register_v1`). Its sole job is column renaming and minor structural tidying:

- All column names are normalised to `snake_case`.
- `book_number` and `page_number` are extracted from `Page_ID` using substring positions (characters 2-6 and 7-9 respectively) as a legacy fallback.
- The `smea_result_5` typo present in the raw export is aliased to `smear_result_5` at this stage.
- No rows are filtered; all scanned records - including discarded, unverified, and duplicate entries - pass through.
- Every select-one field is exported as **individual boolean columns** per option (e.g. `participant_sex_male`, `participant_sex_female`), reflecting ScanForm's materialised export format.

i Note

Why keep everything at the base layer? Cleaning decisions (filtering discards, handling invalids) must happen downstream so that checks can audit the full raw universe of records.

11.2 Pre-Clean - Two models

11.2.1 `mw_tb_pre_clean__tb_reg_raw`

The first pre-clean model applies **row-level exclusion filters** to the base table:

- Removes rows where `discard_row = true`
- Removes pages where `discard_page = true`
- Excludes records where `tb_reg_date` is the blank template string `'__/__/20__'`
- Excludes records where `tb_reg_number` equals the blank template `'MTB-____-____'`

A `record_identifier` surrogate key is constructed as `Page_ID || '-' || Record_number` to uniquely identify each patient row across the dataset.

All other columns remain as raw text strings - **no type casting, no value mapping** - preserving the data exactly as OCR produced it so that the checks model can audit raw values.

11.2.2 `mw_tb_pre_clean__tb_reg`

The second pre-clean model applies **field-level transformations** to the filtered raw table:

Table 11.1: Pre-clean transformations applied per field (mw tb pre clean tb reg)

| Output Field | Transformation | Detail |
|---|-------------------------------------|--|
| age years | Unit-aware numeric conversion | If age unit = Y: age as float. If age unit = M: age / 12. Yields decimal years. |
| sex | Melt boolean columns | participant sex male / participant sex female to single 'Male' / 'Female' value |
| risk group | Melt boolean columns | 8 individual risk columns to single label (Miner, Ex-Miner, Mining Community, HCW, Prisoner, HH Contact, Migrant, Other) |
| referral | Melt boolean columns | 7 referral columns to Community, NCD, Private, Walk In, OPD, Ward, ART |
| treatment regimen | Melt boolean columns | 4 regimen columns to 2RHZE/4RH, BPAlM, BPal, Other Regimen |
| treatment regimen other | pretty_text() | Removes underscores from OCR-produced spacing artefacts |
| disease site | Melt boolean columns | Pulmonary / Extra-Pulmonary |
| treatment history | Melt boolean columns | New, Relap, RALF, Fail, Other, Unknown |
| how diagnosed | Melt boolean columns | Bacteriologically / Clinically |
| smear result (x4) | Digit + plus-sign combination logic | 0 to Negative; 1/2/3 + oval to '1+'/'2+'/'3+'; 1/2/3 without oval to exact count; 4-9 to exact count; else Invalid |
| lam/xpert/culture/xray/dst results | Melt boolean columns | Two- or three-option ovals collapsed to single label string |
| dst result | Melt boolean columns | RIF Resistant Detected / RIF Resistant Not Detected / RIF Resistant Indeterminate |
| treat outcome | Melt boolean columns | Cured, Completed, Dead, Treatment Failure, Lost to Follow-up, Not Evaluated |
| adherence support | Melt boolean columns | Guardian, Health Surveillance Assistant, Volunteer, Healthcare Worker |
| hiv status | Melt boolean columns | Positive / Negative / Unknown |
| arv status | Melt boolean columns | Before TB Tx / While on TB Tx / Not While on TB Tx |
| arv id | Null-blank replacement | ' ' (13 underscores) to NULL |
| art reg id | Null-blank replacement | 'ART- - - ' to NULL |
| participant reg id | Null-blank replacement | ' - - ' to NULL |
| tb reg number | Null-blank replacement | 'MTB- - - ' to NULL |
| All date fields | convert_to_date() | String to DATE type using dd/mm/yy or dd/mm/yyyy format macro |
| enrolled prp / cpt status / second line | Melt boolean columns | Yes / No |
| ptld how diagnosed | Melt boolean columns | Spirometry/LFT, Persistent Respiratory Symptoms, Abnormalities X-ray, Reduced Exercise |
| prp outcome | Melt boolean columns | Completed, Died, Lost to Follow-up, Active TB, Excluded |

💡 Smear result combination logic

Smear microscopy is recorded on paper as two complementary fields: a **digit box** (0-9 bacilli count) and a **plus oval** (to indicate a grade). The pipeline combines them as follows:

| Digit | Plus oval marked? | Clean value |
|-------|-------------------|-------------|
| 0 | - | Negative |
| 1 | No | 1 |
| 1 | Yes | 1+ |
| 2 | No | 2 |
| 2 | Yes | 2+ |
| 3 | No | 3 |
| 3 | Yes | 3+ |

| | | |
|-------|---|-------------|
| 4-9 | - | Exact count |
| Other | - | Invalid |

This logic is applied identically to all four smear timepoints (baseline, 2-month, 5-month, 6-month).

11.3 Clean - Two models

11.3.1 mw_tb_clean__tb_reg_checks

The checks model runs **data quality assertions** against `mw_tb_pre_clean__tb_reg_raw` (raw strings, before type casting) and unions them into a single checks table. Each row in the output represents one triggered check for one record.

Table 11.3: Data quality checks (mw tb clean tb reg checks)

| Check | Input Columns | Severity | Tag | Description |
|---------------------------------------|---|----------|--------------------------------|---|
| convert to int | participant age, smear result (x4) | Warning | invalid-number-check | Flags records where age or smear digit boxes cannot be cast to integer |
| convert to date (Warning) | lam date, smear date, smear date 2, smear date 5, xpert date, culture date, xray date, dst date, cpt start date | Warning | invalid-date-check | Flags unparseable dates for optional date fields |
| convert to date (Error) | tb reg date, outcome date, smear date 6 | Error | invalid-date-check | Flags unparseable dates for key date fields; these records are excluded from clean data |
| batch check melt columns | All select-one fields (22 groups) | Warning | multiple-answer-selected-check | Flags records where more than one oval was marked for a single-select question |
| drug resistant standard regimen check | dst result dst res + treatment regimen regimen stan | Error | clinical-logic-check | Flags records where DST is RIF Resistant Detected but the standard 2RHZE/4RH regimen is recorded - a clinical contradiction |

! Important

Error vs Warning

- **Error** - the record is **excluded** from `mw_tb_clean__tb_reg` and all downstream models. Errors indicate data that cannot be meaningfully interpreted (e.g. an unparseable registration date).
- **Warning** - the record **passes through** to the clean table but the issue is flagged for review in the checks dashboard (e.g. a multiple-selection on a single-select oval field).

11.3.2 mw_tb_clean__tb_reg

The clean table selects all transformed fields from `mw_tb_pre_clean__tb_reg` and **filters out** any `record_identifier` that triggered an **Error**-severity check. The resulting dataset contains only records that are structurally valid and interpretable, with all fields cast to appropriate types and all select-one columns collapsed to clean label strings.

11.4 Intermediary - Two models

11.4.1 mw_tb_intermediary__tb_aggregated

Produces a **single aggregate summary row** across all clean records. Each patient record is first scored with binary flags, then summed:

Table 11.4: Aggregated summary metrics (mw tb intermediary tb aggregated)

| Metric | Aggregation Logic |
|--------------------------|---|
| total treated | Count of records where tb_reg_date is not NULL |
| average age | Mean of age_years across all records |
| total female | Count where sex = 'Female' |
| total male | Count where sex = 'Male' |
| total standard regimen | Count where treatment_regimen = '2RHZE/4RH' |
| total other regimen | Count where treatment_regimen = 'Other Regimen' |
| total pulmonary tb | Count where disease_site = 'Pulmonary' |
| total previously treated | Count where treatment_history IN ('Relap', 'RALF', 'Fail', 'Other') |
| total resistant tests | Count where dst_date is not NULL |
| total resistant tb | Count where dst_result = 'RIF Resistant Detected' |
| total positive outcome | Count where treat_outcome IN ('Cured', 'Completed') |
| total hiv positive | Count where hiv_status = 'Positive' |
| total enrolled prp | Count where enrolled_prp = 'Yes' |
| total completed prp | Count where prp_outcome = 'Completed' |

11.4.2 mw_tb_intermediary__checks_with_locations

Joins `mw_tb_clean__tb_reg_checks` with `mw_tb_intermediary__books_with_locations` on `book_number` to **enrich check records with facility/location metadata**. This powers location-disaggregated DQA dashboards where data managers can see which site generated the most errors or warnings.

11.5 Refinery - Two models

Refinery models are the **final delivery layer** consumed by dashboards and external sinks (e.g. DHIS2, Metabase). They mirror the clean and checks content with any final adjustments needed for downstream consumers.

11.5.1 mw_tb_refinery__tb_reg

A pass-through of all columns from `mw_tb_clean__tb_reg`. No further transformation is applied; the refinery layer simply exposes the clean patient-level dataset in the `marts` schema where it is accessible to dashboard tooling.

11.5.2 mw_tb_refinery__tb_reg_checks

Selects from mw_tb_clean__tb_reg_checks and applies one additional filter:

- **Excludes rows where implementation_error = true** - implementation errors reflect paper form issues (e.g. duplicate rows printed in error) rather than genuine data quality problems, and should not appear in operational DQA reports shown to facility staff.

12 Key Design Decisions

Table 12.1: Key pipeline design decisions

| Decision | Rationale |
|--|--|
| Raw export kept as text strings throughout pre clean raw record identifier = Page ID + Record number Errors exclude from clean; Warnings do not | Allows checks to inspect the actual OCR output (e.g. ' / /20 ') before any casting masks the raw value Provides a stable surrogate key linking checks back to the specific patient row and page scan Errors (e.g. unparseable registration date, clinical contradiction) make a record unanalysable. Warnings flag issues for review without losing the record entirely |
| Blank template strings to NULL | ScanForm exports unfilled structured fields as the template pattern (e.g. 'MTB- - '). These are replaced with NULL so that downstream aggregations and joins treat them as missing rather than a valid value |
| Age normalised to decimal years | Infants recorded in months are divided by 12 so that age years is a single comparable numeric field across all patients |
| Smear result combines digit + plus oval implementation error filtered from refinery checks book number extracted from Page ID substring | The paper form encodes grade information across two separate OCR fields. The pipeline merges them into a single interpretable clinical value (e.g. '2+') Facility staff reviewing DQA reports should not see errors caused by form printing issues outside their control Legacy handling: in the absence of a separate book metadata export, position 2-6 of Page ID encodes the book identifier |

13 Model Inventory

Table 13.1: All pipeline models (9 models across 5 layers)

| Model | Layer | Reads From | Output Granularity | Notes |
|--|---------------------|--|------------------------------------|--|
| mw tb base mat pip MW TB Register v1 | Base | scanform materialized exports | All scanned rows | Column rename only; no filtering |
| mw tb pre clean tb reg raw | Pre-Clean | base mat pip MW TB Register v1 | Excludes discards | Row filters + record identifier; all values remain as raw text |
| mw tb pre clean tb reg | Pre-Clean | pre clean tb reg raw | Same as raw | Type casts, melt, smear combination, blank to NULL |
| mw tb clean tb reg checks | Clean | pre clean tb reg raw | One row per check triggered | Union of 5 check types; Error and Warning severity |
| mw tb clean tb reg | Clean | pre clean tb reg | Excludes Error records | Final clean patient-level table |
| mw tb intermediary tb aggregated | Intermediary | clean tb reg | 1 summary row | 14 aggregate metrics for dashboard KPI tiles |
| mw tb intermediary checks with locations | Intermediary | clean tb reg checks + books with locations | One row per check (with location) | Joins checks to facility metadata |
| mw tb refinery tb reg | Refinery | clean tb reg | Same as clean | Pass-through; exposed in marts schema |
| mw tb refinery tb reg checks | Refinery | clean tb reg checks | Excludes implementation error rows | DQA checks surfaced to facility staff |

14 Variables

15 Overview

The **TB Facility Register** captures data for up to **5 patients per page**. All fields are physically printed on the paper form. The sections below document every variable, its input type, whether it is required, any fill condition, validation constraints, and the OCR model used for automated recognition.

i Note

Reading the tables

- **Required (Yes)** - field must be completed for every patient row.
- **Conditional (Yes)** - field is present on paper but should only be filled when a specific preceding answer is given (see the *Fill Condition* column).
- **OCR Model** - the AI model ScanForm applies to recognise the handwritten or marked value. Fields with - are free-text handwriting zones not processed by OCR.

16 Field Types at a Glance

Table 16.1: Field types across all 60 variables

| Input Type | Count | Share |
|--------------------|-------|-------|
| Oval bubbles | 22 | 37% |
| Digit boxes | 15 | 25% |
| Digit box | 5 | 8% |
| Handwriting | 5 | 8% |
| Checkbox | 4 | 7% |
| Oval bubble | 4 | 7% |
| Letter/digit boxes | 3 | 5% |
| Digit/letter boxes | 2 | 3% |

17 Variables by Section

Table 17.1: Variable count per section

| Section | Variables |
|---------------------|-----------|
| Patient Identity | 3 |
| Registration | 3 |
| Demographics | 3 |
| Risk Group | 2 |
| Referral | 1 |
| Treatment | 5 |
| Initial Lab Results | 9 |
| Follow-up Smears | 9 |
| Imaging and DST | 4 |
| Treatment Outcome | 4 |
| HIV and ART | 7 |
| Post-TB and PRP | 3 |
| Other | 1 |
| Administrative | 6 |

18 Full Variable Reference

18.1 Patient Identity

Table 18.1: Section: Patient Identity

| Variable | Label | Input Type | Re-quired | Fill Condition | OCR Model | Notes |
|--------------|--------------|-------------|-----------|----------------|-----------|-----------|
| full name | Full Name | Handwriting | No | - | - | Free text |
| full address | Address | Handwriting | No | - | - | Free text |
| phone number | Phone Number | Handwriting | No | - | - | Free text |

18.2 Registration

Table 18.2: Section: Registration

| Variable | Label | Input Type | Re-quired | Fill Condition | OCR Model | Notes |
|--------------------|-----------------------------|--------------------|-----------|----------------|--|------------------------------------|
| tb reg date | Registration Date | Digit boxes | Yes | - | date dd x2, date mm x2, int x2, date nearby yy x2 | dd/mm/yyyy; century 20 pre-printed |
| participant reg id | Participant Registration ID | Digit/letter boxes | No | - | int x8, string restricted BFGHJPVWXZ x2 | Copied from TB Screening Form |
| tb reg number | TB Registration Number | Digit/letter boxes | Yes | - | string x3, int x5, record int x3, mtb check symbols x2 | Prefix MTB-pre-printed |

Pre-printed characters explained

- `tb_reg_date` format `[] [] / [] [] / [2] [0] [] []`: the digits 2 and 0 are **pre-printed** on the paper so staff write only the last two digits of the year.
- `tb_reg_number` format `[M] [T] [B] - [] [] [] [] - [] [] [] - [] []`: the letters M, T, B are **pre-printed**; staff complete the numeric segments only.

18.3 Demographics

Table 18.3: Section: Demographics

| Variable | Label | Input Type | Re-quired | Fill Condition | OCR Model | Notes |
|----------------------|-------|--------------|-----------|----------------|--------------------|----------------------|
| participant age | Age | Digit boxes | Yes | - | int | Enter months if <2 y |
| participant age unit | Y/M | Digit box | Yes | - | age unit check | Y = years, M = mon |
| participant sex | Sex | Oval bubbles | Yes | - | select one or zero | Male / Female |

💡 Age unit field (participant_age_unit)

Write **Y** for years or **M** for months. If the patient is under 2 years old, record age in months and mark **M**.

18.4 Risk Group

Table 18.4: Section: Risk Group

| Variable | Label | Input Type | Re-quired | Fill Condition | OCR Model | Notes |
|------------------------|-------------------------|--------------|-----------|----------------|--------------------|--|
| participant risk | Occupation / Risk Group | Oval bubbles | Yes | - | select one or zero | Miner, Ex-Miner, Mining Community, HCW, Prisoner, HH Contact, Migrant, O |
| participant risk other | If Other, Specify | Handwriting | No | - | - | Free text |

18.5 Referral

Table 18.5: Section: Referral

| Variable | Label | Input Type | Re-quired | Fill Condition | OCR Model | Notes |
|-----------------|----------|--------------|-----------|----------------|--------------------|--|
| referral source | Referral | Oval bubbles | Yes | - | select one or zero | Community, NCD, Private, Walk in, OP Ward, ART |

18.6 Treatment

Table 18.6: Section: Treatment

| Variable | Label | Input Type | Re-quired | Fill Condition | OCR Model | Notes |
|-------------------------|-------------------|--------------------|-----------|-----------------------------------|--------------------|--|
| treatment regimen | Treatment Regimen | Oval bubbles | Yes | - | select one or zero | 2RHZE/4RH, BPalm BPalm, Other |
| treatment regimen other | Other Regimen | Letter/digit boxes | No | treatment regimen = regimen other | string int | Fill only if Other regimen selected |
| disease site | Site of Disease | Oval bubbles | Yes | - | select one or zero | Pulmonary, Extra-Pulmonary |
| tb history | Treatment History | Oval bubbles | Yes | - | select one or zero | New, Relapse, RALF |
| how diagnosed | Bact/Cl Diagnosed | Oval bubbles | Yes | - | select one or zero | Fail, Other, Unknown Bacteriologically, Clinically |

i Treatment History codes

| Code | Meaning |
|---------|---|
| New | New case |
| Relapse | Previously treated - cured or completed |

| | |
|----------------|------------------------------------|
| RALF | Return after lost to follow-up |
| Fail | Treatment after treatment failure |
| Other | Previous treatment outcome unknown |
| Unknown | Treatment history unknown |

18.7 Initial Lab Results

Table 18.8: Section: Initial Lab Results

| Variable | Label | Input Type | Re-quired | Fill Condition | OCR Model | Notes |
|----------------|----------------|--------------|-----------|----------------|---|---|
| lam date | LAM Date | Digit boxes | No | - | date dd x2, date mm x2, date nearby yy x2 | dd/mm/yy |
| lam result | LAM Result | Oval bubbles | No | - | select one or zero | MTB Detected, MTB Not Detected |
| smear date | Smear Date | Digit boxes | No | - | date dd x2, date mm x2, date nearby yy x2 | dd/mm/yy |
| smear result | Smear Result | Digit box | No | - | int | Count; see smear plu for grade |
| smear plus | Smear Plus | Oval bubble | No | - | select one or zero | Mark + if 1+, 2+ or 3+; leave blank if 0 or exact count |
| xpert date | Xpert Date | Digit boxes | No | - | date dd x2, date mm x2, date nearby yy x2 | dd/mm/yy |
| xpert result | Xpert Result | Oval bubbles | No | - | select one or zero | MTB Detected, MTB Not Detected |
| culture date | Culture Date | Digit boxes | No | - | date dd x2, date mm x2, date nearby yy x2 | dd/mm/yy |
| culture result | Culture Result | Oval bubbles | No | - | select one or zero | MTB Detected, MTB Not Detected |

Smear result entry

`smear_result` accepts a digit (0-9) representing the bacilli count. The accompanying `smear_plus` oval should be crossed only when the grade is 1+, 2+, or 3+. Leave the oval blank when recording an exact count or a result of 0.

18.8 Follow-up Smears

Table 18.9: Section: Follow-up Smears

| Variable | Label | Input Type | Re-quired | Fill Condition | OCR Model | Notes |
|----------------|------------------------|-------------|-----------|----------------|---|------------------------|
| smear date 2 | Smear Date - 2 month | Digit boxes | No | - | date dd x2, date mm x2, date nearby yy x2 | dd/mm/yy |
| smear result 2 | Smear Result - 2 month | Digit box | No | - | int | - |
| smear plus 2 | Smear Plus - 2 month | Oval bubble | No | - | select one or zero | Mark + if 1+, 2+ or 3+ |
| smear date 5 | Smear Date - 5 month | Digit boxes | No | - | date dd x2, date mm x2, date nearby yy x2 | dd/mm/yy |

Table 18.9: Section: Follow-up Smears (*continued*)

| Variable | Label | Input Type | Re-quired | Fill Condition | OCR Model | Notes |
|------------------------------|--|----------------------------|-----------|----------------|---|--|
| smea result 5 | Smear Result - 5 month | Digit box | No | - | int | Typo in variable name smea result 5 |
| smear plus 5 smear date 6 | Smear Plus - 5 month Smear Date - 6 month | Oval bubble Digit boxes | No No | - - | select one or zero date dd x2, date mm x2, date nearby yy x2 | Mark + if 1+, 2+ or dd/mm/yy |
| smear result 6 | Smear Result - 6 month | Digit box | No | - | int | - |
| smear plus 6 | Smear Plus - 6 month | Oval bubble | No | - | select one or zero | Mark + if 1+, 2+ or |

 Warning

smea_result_5 contains a **typo** in the variable name (missing letter r). Data pipelines must reference this exact name.

18.9 Imaging and DST

Table 18.10: Section: Imaging and DST

| Variable | Label | Input Type | Re-quired | Fill Condition | OCR Model | Notes |
|-------------|-------------|--------------|-----------|----------------|---|--|
| xray date | XRay Date | Digit boxes | No | - | date dd x2, date mm x2, date nearby yy x2 | dd/mm/yy |
| xray result | XRay Result | Oval bubbles | No | - | select one or zero | MTB Detected, MTB Not Detected |
| dst date | DST Date | Digit boxes | No | - | date dd x2, date mm x2, date nearby yy x2 | dd/mm/yy |
| dst result | DST Result | Oval bubbles | No | - | select one or zero | RIF Resistant Detect RIF Resistant Not Detected, RIF Resist Indeterminate |

18.10 Treatment Outcome

Table 18.11: Section: Treatment Outcome

| Variable | Label | Input Type | Re-quired | Fill Condition | OCR Model | Notes |
|----------------------------------|--|------------------------------|-----------|----------------|---|--|
| treat outcome | Treatment Outcome | Oval bubbles | No | - | select one or zero | Cured, Treatment Completed (No smea Dead, Treatment Failure, Lost to Follow-up, Not Evaluated |
| outcome date | Outcome Date | Digit boxes | No | - | date dd x2, date mm x2, date nearby yy x2 | dd/mm/yy |
| second line adherence support | Moved to 2nd Line Tx DOT and Adherence Support | Oval bubbles Oval bubbles | No No | - - | select one or zero select one or zero | Yes, No GUA, HSA, VOL, H |

18.11 HIV and ART

Table 18.12: Section: HIV and ART

| Variable | Label | Input Type | Re-quired | Fill Condition | OCR Model | Notes |
|----------------|---------------------|--------------------|-----------|---------------------------------------|--|--|
| hiv status | HIV Status | Oval bubbles | Yes | - | select one or zero | Pos, Neg, Unk |
| hiv test time | HIV Test Time | Oval bubbles | No | hiv status = positive or negative | select one or zero | Before TB Reg, After TB Reg |
| arv status | ARV Start Time | Oval bubbles | No | hiv status = positive | select one or zero | Before TB Tx, While on TB Tx, Not While on TB Tx |
| arv id | ARV No. | Letter/digit boxes | No | arv status = before tb or while on tb | string int | - |
| art reg id | ART Registration ID | Letter/digit boxes | No | hiv status = positive | string x3, int x8, string restricted BFGHJPVWXZ x2 | Prefix ART- pre-printed |
| cpt status | CPT Status | Oval bubbles | No | hiv status = positive | select one or zero | Yes, No |
| cpt start date | CPT Start Date | Digit boxes | No | cpt status = yes | date dd x2, date mm x2, date nearby yy x2 | dd/mm/yy |

! HIV-related conditional logic

Several fields are only relevant when HIV is confirmed positive. Complete them **only** when `hiv_status = Pos`:

- `arv_status` - ARV Start Time
- `art_reg_id` - ART Registration ID (ART- pre-printed)
- `cpt_status` - CPT Status
- `cpt_start_date` - CPT Start Date (*additionally requires `cpt_status = Yes`*)
- `arv_id` - ARV No. (*additionally requires `arv_status = Before TB Tx or While on TB Tx`*)

`hiv_test_time` should be completed when HIV status is **Pos** or **Neg**.

18.12 Post-TB and PRP

Table 18.13: Section: Post-TB and PRP

| Variable | Label | Input Type | Re-quired | Fill Condition | OCR Model | Notes |
|--------------------|-----------------------|--------------|-----------|------------------------------------|--------------------|---|
| enrolled prp | Enrolled into PRP | Oval bubbles | No | treat outcome = cured or completed | select one or zero | Yes, No |
| ptld how diagnosed | PTLD Diagnosis Method | Oval bubbles | No | enrolled prp = yes | select one or zero | Spirometry/LFT, Persistent Respiratory Symptoms, Abnormalities Xray, Reduced Exercise |
| prp outcome | PRP Outcome | Oval bubbles | No | - | select one or zero | Completed, Died, Lost to Follow-up, Active TB, Excluded |

i PRP enrolment logic

enrolled_prp is filled only when treat_outcome is **Cured** or **Completed**. If enrolled, ptld_how_diagnosed records the diagnostic method. prp_outcome is completed for all PRP patients regardless of enrolment route.

18.13 Other

Table 18.14: Section: Other

| Variable | Label | Input Type | Re-quired | Fill Condition | OCR Model | Notes |
|----------|----------|-------------|-----------|----------------|-----------|-----------|
| comments | Comments | Handwriting | No | - | - | Free text |

18.14 Administrative

Table 18.15: Section: Administrative

| Variable | Label | Input Type | Re-quired | Fill Condition | OCR Model | Notes |
|----------------------|-----------------------------|-------------|-----------|----------------|-----------|---|
| book number | Book Number | Digit boxes | No | - | - | Page-level field |
| page number | Page Number | Digit boxes | No | - | - | Page-level field |
| photo taken | Photo Taken | Checkbox | No | - | - | Confirms scan photo captured |
| discard page | Discard Left and Right Page | Checkbox | No | - | - | Marks entire page as discarded |
| discard row | Discard Row | Checkbox | No | - | - | Marks individual patient row as discarded |
| implementation error | Implementation Error | Checkbox | No | - | - | Flags a data entry error |

i Administrative field notes

- **Book Number** (5 boxes) and **Page Number** (3 boxes) are page-level identifiers shared by all 5 patient rows on the page.
- **Photo Taken**, **Discard Left and Right Page**, **Discard Row**, and **Implementation Error** are trigger checkboxes - they carry no free-text value and are not processed by OCR.

19 Conditional Logic Map

Table 19.1: Conditional fields (9 of 60 variables) - fill only when condition is met

| Variable | Label | Section | Fill Condition |
|-------------------------|-----------------------|-----------------|---------------------------------------|
| treatment regimen other | Other Regimen | Treatment | treatment regimen = regimen other |
| hiv test time | HIV Test Time | HIV and ART | hiv status = positive or negative |
| arv status | ARV Start Time | HIV and ART | hiv status = positive |
| arv id | ARV No. | HIV and ART | arv status = before tb or while on tb |
| art reg id | ART Registration ID | HIV and ART | hiv status = positive |
| cpt status | CPT Status | HIV and ART | hiv status = positive |
| cpt start date | CPT Start Date | HIV and ART | cpt status = yes |
| enrolled prp | Enrolled into PRP | Post-TB and PRP | treat outcome = cured or completed |
| ptld how diagnosed | PTLD Diagnosis Method | Post-TB and PRP | enrolled prp = yes |

20 OCR Models Used

Table 20.1: OCR models deployed in this form (each box may use a distinct model)

| OCR Model | Fields Using | Category |
|---------------------------------|--------------|----------------------------|
| date dd x2 | 12 | <i>Date component</i> |
| date mm x2 | 12 | <i>Date component</i> |
| date nearby yy x2 | 12 | <i>Date component</i> |
| int | 5 | <i>General</i> |
| string int | 2 | <i>General</i> |
| int x8 | 2 | <i>Other</i> |
| string x3 | 2 | <i>Other</i> |
| age unit check | 1 | <i>Other</i> |
| int x2 | 1 | <i>Other</i> |
| int x5 | 1 | <i>Other</i> |
| mtb check symbols x2 | 1 | <i>Other</i> |
| record int x3 | 1 | <i>Other</i> |
| select one or zero | 26 | <i>Oval / select</i> |
| string restricted BFGHJPVWXZ x2 | 2 | <i>Restricted alphabet</i> |

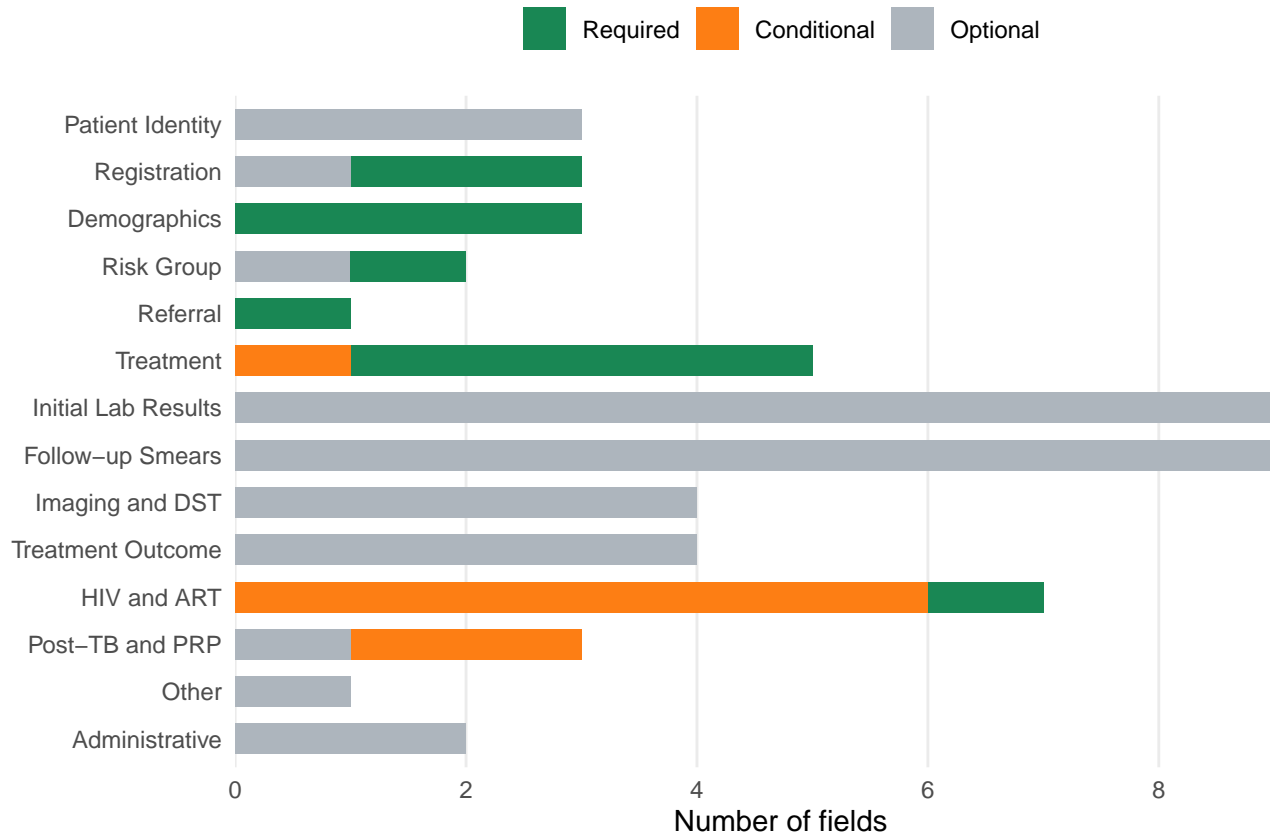
💡 What do these models do?

| Model | What it recognises |
|-------------------------------|---|
| select one or zero | Oval bubbles: detects X or blank; at most one X per question |
| int | Digits 0-9 |
| string | Letters A-Z |
| string int | Letters A-Z and digits 0-9 |
| date dd | Two-digit day (1-31) |
| date mm | Two-digit month (1-12) |
| date nearby yy | Two-digit year near current year (+/-1) |
| string restricted BFGHJPVWXZ | Restricted letter set used in ID codes to avoid OCR confusion |
| age unit check | Custom local model: Y or M only |
| record int, mtb check symbols | Custom local models for TB registration number validation |

21 Required vs Optional Fields

Required, conditional, and optional fields by section

Trigger/administrative checkboxes excluded



22 In-Phone Validation

23 How in-phone validation works

When a field worker photographs a completed form page, ScanForm's OCR engine reads the paper immediately and runs a set of automated checks **before the data is submitted**. If any check fails, the app displays an alert on-screen and prompts the worker to correct the paper form and retake the photo.

! Important

This is fully automated, real-time validation - no human verifier is involved at this stage. The goal is to catch critical errors at the point of data capture, while the form is still in hand and corrections are still possible.

The validation logic is defined in `data_validation.py` using a Python DSL (`ChecksBuilder`). The sections below explain the key concepts, the eligibility rules, and every check applied to this form.

24 Core concepts

Table 24.1: ScanForm validation DSL - key concepts

| Concept | Meaning |
|---|--|
| <code>b.question(name, row, boxals_threshold_for_eligibility=N)</code> | Registers a field for validation. The threshold N is the minimum number of filled boxes required for the field to be considered active (i.e. the worker has started filling it in). |
| <code>b.any_active(checked_groups, checked_groups_threshold=N, discard_criteria=[...])</code> | Defines row eligibility. A patient row is only validated if $\geq N$ of the specified key fields are active AND no discard bubble is marked. This prevents false alerts on genuinely blank rows at the bottom of a page. |
| <code>check boxes enough filled</code> | The field must have at least 1 filled box. Used to enforce required digit-box or date fields. |
| <code>check checkboxes enough answers + check checkboxes not too many answers (paired)</code> | Together these enforce exactly one oval bubble selected - used for mandatory single-select fields. |
| <code>check checkboxes not too many answers (alone)</code> | At most one oval bubble may be selected. The field is optional, but if filled it must not have multiple ovals crossed - used for optional single-select fields. |
| <code>PageNumberBlock(0)</code> | Left page of the physical spread. |
| <code>PageNumberBlock(1)</code> | Right page of the physical spread. |
| <code>Commented-out lines</code> | Fields that exist on the paper form and are exported by OCR, but are not actively validated in-phone. They are available in the dataset but will not trigger an on-screen alert. |

25 Eligibility rules

Before any field-level check is run, the system first determines whether a patient row is **eligible for validation**. This two-stage gate prevents spurious alerts on empty rows.

25.1 Page-level gate

A discard bubble (`discard_page`) is read once at page level. If it is marked, **all 5 patient rows on both pages are skipped** - no field checks are triggered at all.

25.2 Row-level gate

For each of the 5 patient rows, the system checks:

i Note

A row is validated only if:

1. `discard_page` is **not** marked, AND
2. `discard_row` is **not** marked for that row, AND
3. **At least 3** of the following 12 key fields are active (i.e. have ≥ 1 filled box or bubble): `tb_reg_date`, `tb_reg_number`, `participant_age`, `participant_age_unit`, `participant_sex`, `participant_risk`, `referral_source`, `treatment_regimen`, `disease_site`, `tb_history`, `how_diagnosed`, `hiv_status`

The threshold of **3 out of 12** is deliberately generous - it means that even a partially completed row (e.g. only the date, number, and one oval filled) will trigger validation, while a completely blank row at the bottom of the page will not.

Table 25.1: Fields used in row eligibility assessment (≥ 3 of these 12 must be active)

| Variable | Label | Eligibility Threshold | Notes |
|-----------------------------------|-------------------------|-----------------------|-------------------------------|
| <code>tb_reg_date</code> | Registration Date | 2 boxes | At least 2 boxes filled |
| <code>tb_reg_number</code> | TB Registration Number | 3 boxes | At least 3 boxes filled |
| <code>participant_age</code> | Age | 1 (default) | Any fill |
| <code>participant_age_unit</code> | Y/M | 1 (default) | Any fill |
| <code>participant_sex</code> | Sex | 1 (default) | Any oval |
| <code>participant_risk</code> | Occupation / Risk Group | 2 | At least 2 boxes/ovals filled |
| <code>referral_source</code> | Referral | 2 | At least 2 boxes/ovals filled |
| <code>treatment_regimen</code> | Treatment Regimen | 1 (default) | Any oval |
| <code>disease_site</code> | Site of Disease | 1 (default) | Any oval |
| <code>tb_history</code> | Treatment History | 2 | At least 2 boxes/ovals filled |

| | | | |
|---------------|-------------------|-------------|----------|
| how diagnosed | Bact/Cl Diagnosed | 1 (default) | Any oval |
| hiv status | HIV Status | 1 (default) | Any oval |

26 Validated fields and checks

The table below lists every field-level check that runs in-phone. Checks are applied to each of the 5 patient rows independently.

Table 26.1: In-phone checks - all 26 validated fields applied per patient row

| Variable | Label | Page | Check Type | Rule | On Failure |
|----------------------|---------------------------|-------|---------------|----------------|-----------------------------|
| tb reg date | Registration Date | Left | Required fill | >=1 box filled | Field is required |
| tb reg number | TB Registration Number | Left | Required fill | >=1 box filled | Field is required |
| participant age | Age | Left | Required fill | >=1 box filled | Field is required |
| participant age unit | Y/M | Left | Required fill | >=1 box filled | Field is required |
| participant sex | Sex | Left | Exactly one | Exactly 1 oval | Exactly one option required |
| participant risk | Occupation / Risk Group | Left | Exactly one | Exactly 1 oval | Exactly one option required |
| referral source | Referral | Left | Exactly one | Exactly 1 oval | Exactly one option required |
| treatment regimen | Treatment Regimen | Left | Exactly one | Exactly 1 oval | Exactly one option required |
| disease site | Site of Disease | Left | Exactly one | Exactly 1 oval | Exactly one option required |
| tb history | Treatment History | Left | Exactly one | Exactly 1 oval | Exactly one option required |
| how diagnosed | Bact/Cl Diagnosed | Left | Exactly one | Exactly 1 oval | Exactly one option required |
| lam result | LAM Result | Left | Max one | <=1 oval | No more than one option |
| xpert result | Xpert Result | Left | Max one | <=1 oval | No more than one option |
| culture result | Culture Result | Left | Max one | <=1 oval | No more than one option |
| xray result | XRay Result | Right | Max one | <=1 oval | No more than one option |
| dst result | DST Result | Right | Max one | <=1 oval | No more than one option |
| second line | Moved to 2nd Line Tx | Right | Max one | <=1 oval | No more than one option |
| treat outcome | Treatment Outcome | Right | Max one | <=1 oval | No more than one option |
| adherence support | DOT and Adherence Support | Right | Max one | <=1 oval | No more than one option |
| hiv status | HIV Status | Right | Exactly one | Exactly 1 oval | Exactly one option required |
| hiv test time | HIV Test Time | Right | Max one | <=1 oval | No more than one option |
| arv status | ARV Start Time | Right | Max one | <=1 oval | No more than one option |
| cpt status | CPT Status | Right | Max one | <=1 oval | No more than one option |
| enrolled prp | Enrolled into PRP | Right | Max one | <=1 oval | No more than one option |
| ptld how diagnosed | PTLD Diagnosis Method | Right | Max one | <=1 oval | No more than one option |
| prp outcome | PRP Outcome | Right | Max one | <=1 oval | No more than one option |

27 Check types explained

Table 27.1: Three check types deployed in this form

| Check Type | Macro Name | Fields Using | Description |
|----------------------|--|--------------|--|
| Required fill | <code>enough filled</code> | 4 | The field must have at least 1 filled box. Applies to digit-box and date fields that are mandatory. If the worker left the field blank on an otherwise active row, an alert fires immediately. |
| Exactly one | <code>enough answers + not too many</code> | 9 | Both a lower and upper bound on the number of crossed ovals. Used for mandatory single-select questions. Fires if the field is blank (no oval crossed) OR if more than one oval is crossed. |
| Max one | <code>not too many</code> | 13 | Upper bound only. The field is optional, so leaving it blank is acceptable. However, if the worker crosses two or more ovals, an alert fires. Used for optional single-select questions. |

28 Fields not validated in-phone

Many fields on the form are registered in the OCR pipeline but are **intentionally excluded** from in-phone validation (shown as commented-out lines in `data_validation.py`). These fields are still OCR-processed and exported - they are simply not checked in real time.

i Note

Exclusion from in-phone validation does not mean the data is ignored. These fields are fully exported and subject to **pipeline-level checks** (see the Data Pipeline tab) after submission.

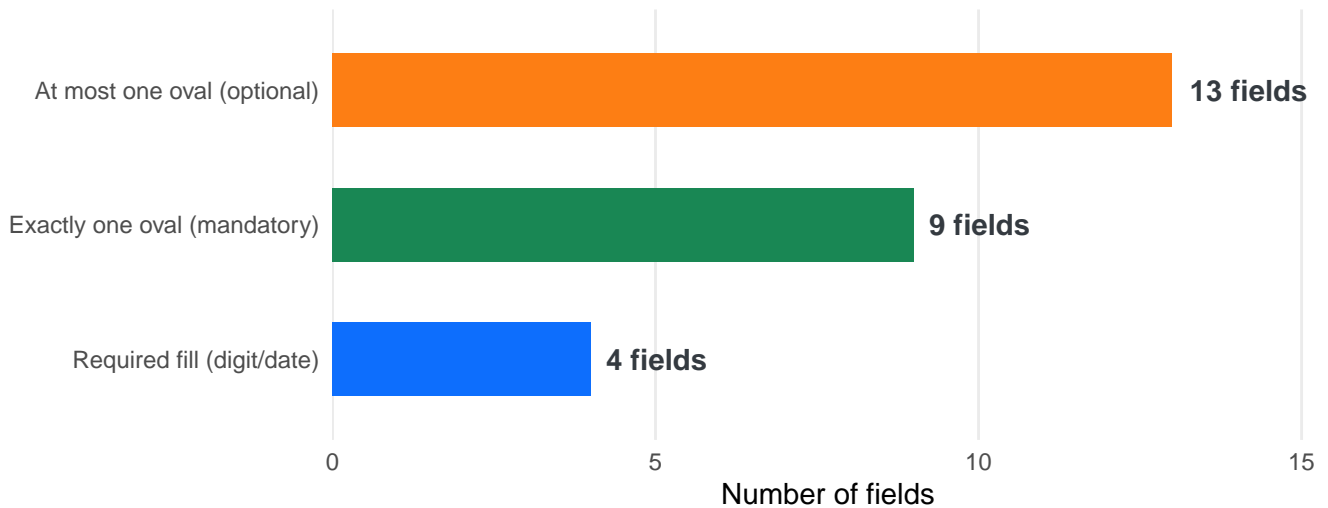
Table 28.1: Fields present on form but excluded from in-phone validation (still OCR-processed and exported)

| Section | Fields / Ovals Not Validated In-Phone |
|---|---|
| Patient identity | participant reg id |
| Demographics | participant sex (individual ovals: male, female) |
| Risk group | All 8 individual risk ovals (miners, ex miners, mining community, health care workers, prisoner, hh contact, migrant, risk other) |
| Referral | All 7 individual referral ovals (community, ncd, private, walk in, opd, ward, art referral) |
| Treatment | All 4 regimen ovals; treatment regimen other; all 2 disease site ovals; all 6 tb history ovals; all 2 how diagnosed ovals |
| Initial lab dates | lam date, smear date, smear result, smear plus, xpert date, culture date |
| Initial lab results (individual ovals) | lam result ovals, xpert result ovals, culture result ovals |
| Follow-up smears | smear date 2/5/6, smear result 2/5/6, smear plus 2/5/6 (all fields and ovals) |
| Imaging & DST | xray date, xray result ovals, dst date, dst result ovals |
| Outcome | outcome date; second line ovals; treat outcome ovals; adherence support ovals |
| HIV / ART | hiv status ovals; hiv test time ovals; arv status ovals; arv id; art reg id; cpt status ovals; cpt start date |
| PRP / PTLD | enrolled prp ovals; ptld how diagnosed ovals; prp outcome ovals |
| Administrative | page number, book number, discard row (checkbox), implementation error, photo taken |

29 Validation summary

In-phone checks by type

Applied to each of the 5 patient rows; 26 unique fields checked per row



💡 Tip

26 fields are validated per patient row. With 5 rows per page, up to **130 individual checks** may fire on a single page scan. The eligibility gate (≥ 3 of 12 key fields active) ensures that only rows where a patient has genuinely been started trigger any checks.

30 Data Quality Checks

31 Overview

After scanned forms are submitted and processed through the OCR pipeline, a second layer of automated data quality checks runs **server-side** against the full dataset. These checks are defined in `mw_tb_clean__tb_reg_checks.sql` and execute as part of the dbt pipeline on every data refresh.

i Note

How this differs from in-phone validation

| Aspect | In-Phone Validation | Pipeline DQA Checks |
|------------------------------|-------------------------------------|---------------------------------------|
| When | At photo capture, before submission | After submission, on the full dataset |
| Who sees it | Field worker, immediately | Data manager, in review dashboard |
| Can prevent bad data? | Yes - worker corrects and retakes | No - data already submitted |
| Scope | 26 fields, 5 rows per page | All fields, all records |
| Triggered by | ScanForm mobile app | dbt pipeline run |

Each check produces **one row per failure**. Records that pass all checks produce no rows. The output table (`mw_tb_clean__tb_reg_checks`) is the single source of truth for data quality review and is consumed by the DQA dashboard.

32 Severity levels

Table 32.1: Check severity levels

| Severity | Effect on Record | Typical Use Case |
|----------------|---|---|
| Error | Record is flagged AND excluded from the clean dataset | Data that cannot be meaningfully interpreted downstream (e.g. unparseable registration date, clinical contradiction) |
| Warning | Record is flagged but retained in the clean dataset | Data issues that should be reviewed but do not invalidate the entire record (e.g. unparseable optional date, multiple ovals marked) |

33 Check catalogue

Five check blocks are defined, unioned into a single output table. Each is described below.

33.1 Integer conversion checks

- **CTE:** `convert_to_int_checks`
- **Macro:** `batch_check_convert_to_int`
- **Severity:** Warning
- **Tag:** `invalid-number-check`

Checks that OCR-produced text values for digit-box fields can be parsed as integers. A Warning row is generated for each field on each record where the raw OCR text cannot be converted.

Table 33.1: Integer conversion checks (Warning) - 5 fields

| Variable | Label | Expected Range | Notes |
|------------------------------|-------------------------|------------------|--|
| <code>participant age</code> | Age | Positive integer | Age in years or months depending on age unit |
| <code>smear result</code> | Smear Result (baseline) | 0-9 | Single digit count; combined with smear plus oval downstream |
| <code>smear result 2</code> | Smear Result - 2 month | 0-9 | Follow-up smear |
| <code>smear result 5</code> | Smear Result - 5 month | 0-9 | Follow-up smear; note source variable has typo smea result 5 |
| <code>smear result 6</code> | Smear Result - 6 month | 0-9 | Follow-up smear |

33.2 Date conversion checks - Warnings

- **CTE:** `convert_to_date_checks`
- **Macro:** `batch_check_convert_to_date`
- **Severity:** Warning
- **Tag:** `invalid-date-check`

Checks that OCR text for **optional** date fields can be parsed as a valid date. Records with unparseable optional dates are flagged but retained in the clean dataset.

Table 33.2: Date conversion checks - Warning (9 fields, optional dates)

| Variable | Label | Expected Format |
|---------------------------|-----------------------|-----------------|
| <code>lam date</code> | LAM Date | dd/mm/yy |
| <code>smear date</code> | Smear Date (baseline) | dd/mm/yy |
| <code>smear date 2</code> | Smear Date - 2 month | dd/mm/yy |
| <code>smear date 5</code> | Smear Date - 5 month | dd/mm/yy |

| | | |
|----------------|----------------|----------|
| xpert date | Xpert Date | dd/mm/yy |
| culture date | Culture Date | dd/mm/yy |
| xray date | XRay Date | dd/mm/yy |
| dst date | DST Date | dd/mm/yy |
| cpt start date | CPT Start Date | dd/mm/yy |

33.3 Date conversion checks - Errors

- **CTE:** `convert_to_date_checks_errors`
- **Macro:** `batch_check_convert_to_date`
- **Severity:** Error
- **Tag:** `invalid-date-check`

Checks that OCR text for **key** date fields can be parsed as a valid date. Records where these fields are unparseable are **excluded from the clean dataset** entirely, because these dates are essential for record identification or outcome analysis.

Table 33.3: Date conversion checks - Error (3 key fields)

| Variable | Label | Expected Format | Why Error (not Warning) |
|---------------------------|----------------------|-----------------|---|
| <code>tb_reg_date</code> | Registration Date | dd/mm/yyyy | Primary temporal anchor for the patient record; used in all joins, filters and DQA metadata |
| <code>outcome_date</code> | Outcome Date | dd/mm/yy | Core outcome field; an unparseable outcome date makes the treatment outcome unanalysable |
| <code>smear_date_6</code> | Smear Date - 6 month | dd/mm/yy | 6-month smear is the key end-of-treatment bacteriological assessment |

! Important

Impact of date Error checks Any record where `tb_reg_date`, `outcome_date`, or `smear_date_6` cannot be parsed as a valid date will appear in the checks table with `severity = 'Error'` and will be **filtered out** of `mw_tb_clean_tb_reg` and all downstream refinery and intermediary models.

33.4 Multiple-answer selection checks

- **CTE:** `melting_checks`
- **Macro:** `batch_check_melt_columns`
- **Severity:** Warning
- **Tag:** `multiple-answer-selected-check`

ScanForm exports each oval option as a separate boolean column (`true/false`). For single-select questions, exactly one column should be `true`. This check flags records where **more than one option is marked** for the same question - indicating either a scanning error or that the field worker crossed multiple ovals.

A Warning row is generated per affected question per record.

Table 33.4: Multiple-answer selection checks - Warning (22 fields)

| Clean Variable | Label | Options | Raw Boolean Columns Checked |
|--------------------|---------------------------|---------|--|
| sex | Sex | 2 | male, female |
| risk group | Occupation / Risk Group | 8 | miners, prisoner, ex miners, hh contact, mining commun, migrant, health care w, risk other |
| referral | Referral | 7 | community, ncd, private, walk in, opd, ward, art referral |
| treatment regimen | Treatment Regimen | 4 | regimen stan, regimen resi1, regimen resi2, regimen othe |
| disease site | Site of Disease | 2 | pulmonary, extra pulmonary |
| treatment history | Treatment History | 6 | new, relap, ralf, fail, other, unknown |
| how diagnosed | Bact/CI Diagnosed | 2 | dx bacteria, dx clinically |
| lam result | LAM Result | 2 | mtb detected, mtb not detected |
| xpert result | Xpert Result | 2 | mtb detected, mtb not detected |
| culture result | Culture Result | 2 | mtb detected, mtb not detecte |
| xray result | XRay Result | 2 | mtb detected, mtb not detected |
| dst result | DST Result | 3 | dst res, dst non res, dst indeterminate |
| second line | Moved to 2nd Line Tx | 2 | yes, no |
| treat outcome | Treatment Outcome | 6 | cured, completed, dead, failure, lost followup, not evaluated |
| adherence support | DOT and Adherence Support | 4 | guardian, health sur a, volunteer, healthcare w |
| hiv status | HIV Status | 3 | positive, negative, unknown |
| hiv test time | HIV Test Time | 2 | hiv test before , hiv test after t |
| arv status | ARV Start Time | 3 | before tb, while on tb, not on tb |
| cpt status | CPT Status | 2 | yes, no |
| enrolled prp | Enrolled into PRP | 2 | yes, no |
| ptld how diagnosed | PTLD Diagnosis Method | 4 | spiro, res symptom, chest xray, reduced exe |
| prp outcome | PRP Outcome | 5 | prp completed, prp died, prp lost followup, prp active TB, prp excluded |

33.5 Clinical logic check

- **CTE:** `drug_resistant_standard_regimen_check`
- **Type:** Custom SQL
- **Severity:** Error
- **Tag:** `clinical-logic-check`

This check implements a **clinical contradiction rule**: a patient cannot simultaneously have a DST result of RIF Resistant Detected and be prescribed the standard first-line regimen (2RHZE/4RH). If both conditions are true on the same record, the record is flagged as an Error and excluded from the clean dataset.

! Important

Check message: *“Patient marked as drug resistant but prescribed standard treatment regimen”*

Trigger condition: `dst_result_dst_res = 'true' AND treatment_regimen_regimen_stan = 'true'`

Table 33.5: Clinical logic check - Error: drug-resistant result combined with standard first-line regimen

| Field | Label | Trigger Value |
|---------------------------------|------------|------------------------|
| <code>dst result dst res</code> | DST Result | RIF Resistant Detected |

treatment regimen regimen stan

Treatment Regimen

2RHZE/4RH (standard)

34 All checks at a glance

Table 34.1: Complete check inventory - mw tb clean tb reg checks (41 field-check combinations across 5 blocks)

| CTE | Check Type | Tag | Severity | Fields | Fires When |
|---------------------------------------|---------------------------|--------------------------------|----------|--------|--|
| convert to int checks | Integer conversion | invalid-number-check | Warning | 5 | OCR text cannot be parsed as integer |
| convert to date checks | Date conversion | invalid-date-check | Warning | 9 | OCR text cannot be parsed as date (optional fields) |
| convert to date checks errors | Date conversion | invalid-date-check | Error | 3 | OCR text cannot be parsed as date (key fields) |
| melting checks | Multiple-answer selection | multiple-answer-selected-check | Warning | 22 | More than one oval marked for a single-select question |
| drug resistant standard regimen check | Clinical logic | clinical-logic-check | Error | 2 | DST = RIF Resistant AND regimen = 2RHZE/4RH simultaneously |

35 Checks output table structure

Every check block produces rows in a common schema. These are unioned into the final `mw_tb_clean__tb_reg_checks` table.

Table 35.1: Output table schema - mw tb clean tb reg checks (one row per check failure per record)

| Column | Type | Description |
|----------------------|---------|---|
| Entry URL | text | URL uniquely identifying the ScanForm submission |
| Page ID | text | Identifier for the scanned page |
| Submission Date | date | Date the form was submitted to ScanForm |
| tb reg date | date | Patient registration date (parsed); used for temporal reporting |
| book number | text | Register book number extracted from Page ID |
| page number | text | Page number within the book |
| Record number | integer | Patient row number (1-5) within the page |
| record identifier | text | Surrogate key: Page ID + '-' + Record number |
| check message | text | Human-readable description of the specific check failure |
| Submitted by | text | User account that submitted the scan |
| implementation error | boolean | True if this row is also flagged as an implementation error |
| discard page | boolean | True if the page was marked as discarded |
| discard row | boolean | True if the patient row was marked as discarded |
| severity | text | 'Error' or 'Warning' (see severity levels above) |
| tags | array | List of check category tags, e.g. ['invalid-date-check'] |

36 Relationship to in-phone validation

Table 36.1: In-phone validation vs. pipeline DQA checks

| Aspect | In-Phone Validation | Pipeline DQA Checks |
|------------------------------|---|---|
| Timing | Instant, at photo capture | After submission, on next pipeline run |
| Audience | Field worker | Data manager / supervisor |
| Can block submission? | Yes - worker must correct and retake | No - data already in system |
| Fields covered | 26 key fields | 41 fields across 5 check blocks |
| Check types | Required fill, exactly one oval, at most one oval | Integer parse, date parse, multi-select, clinical logic |
| Output | On-screen alert in ScanForm app | Row in mw tb clean tb reg checks table |
| Effect on clean data | Prevents bad data entering | Error severity: excludes record; Warning: flags only |
| Overlap | Checks presence and single-selection only | Checks parseable values and clinical consistency |

Tip

The two validation layers are complementary, not redundant.

In-phone validation catches structural problems (missing required fields, multiple ovals) at the moment of capture - when correction is still possible. Pipeline DQA checks verify that OCR outputs are mathematically and clinically coherent after the fact, and provide the permanent audit trail that data managers use to assess data quality over time.

37 Dashboard Indicators

38 Overview

This page documents every dashboard indicator that can be derived from the TB Facility Register refinery models. Each indicator follows the QED dashboard indicator template: goal, definitions, numerator, denominator, disaggregations, recommended visualisation, and data source.

All indicators draw from one or both of the following refinery tables:

Table 38.1: Refinery models powering the dashboard

| Model | Description | Granularity |
|---|---|--------------------------------------|
| <code>mw tb refinery tb reg</code> | Patient-level clean records, one row per patient per page | One row per patient registration |
| <code>mw tb refinery tb reg checks</code> | Data quality check failures, one row per check triggered | One row per check failure per record |

i Note

Available disaggregation dimensions across all indicators: `book_number` (register book, proxy for facility/cohort), `page_number`, `tb_reg_date` (registration date, for time series), `sex`, `age_years`, `risk_group`, `disease_site`, `treatment_regimen`, `hiv_status`, `treat_outcome`. Geography below book level is not present in the refinery model — facility metadata must be joined from an intermediary locations table.

39 Indicator catalogue

The 25 indicators are organised into the following thematic groups:

- Programme Performance
- Bacteriology & Diagnosis
- Treatment Regimen
- HIV & ART
- Post-TB / PRP
- Data Quality

40 Programme Performance

40.1 Total patients registered

Table 40.1: 1. Total patients registered

| | |
|------------------------|--|
| Goal | Monitor the volume of TB patients being registered at facilities over time. Supports workload planning and programme scale assessment. |
| Numerator | COUNT(*) from mw tb refinery tb reg where tb reg date IS NOT NULL |
| Denominator | N/A - this is an absolute count |
| Disaggregations | book number (facility/cohort), tb reg date (monthly or quarterly time series), sex, disease site |
| Visualization | Line chart - X-axis: tb reg date by month; Y-axis: count of registrations. Add a reference line for target if available. |
| Data Source | mw tb refinery tb reg |

40.2 Treatment success rate

Table 40.2: 2. Treatment success rate

| | |
|----------------------------|---|
| Goal | Assess the proportion of TB patients achieving a successful treatment outcome (Cured or Completed). Primary WHO End TB programme performance indicator. |
| Definition of Terms | Successful outcome: treat outcome IN ('Cured', 'Completed'). Evaluated cohort: all records where treat outcome IS NOT NULL. |
| Numerator | COUNT(*) where treat outcome IN ('Cured', 'Completed') |
| Denominator | COUNT(*) where treat outcome IS NOT NULL |
| Disaggregations | book number, tb reg date (quarterly cohort), sex, age years (grouped: 0-14, 15-24, 25-44, 45-64, 65+), disease site, treatment regimen |
| Visualization | Grouped bar chart - X-axis: time cohort (quarter); Y-axis: percentage. Stack or group by outcome category. Add WHO 90% target line. |
| Data Source | mw tb refinery tb reg |

40.3 Treatment outcome distribution

Table 40.3: 3. Treatment outcome distribution

| | |
|-------------|---|
| Goal | Provide a full breakdown of all treatment outcomes to identify which adverse outcomes (death, failure, lost to follow-up) are most prevalent and where. |
|-------------|---|

| | |
|----------------------------|--|
| Definition of Terms | treat outcome values: 'Cured', 'Completed', 'Dead', 'Treatment Failure', 'Lost to Follow-up', 'Not Evaluated'. Records where treat outcome IS NULL are excluded. |
| Numerator | COUNT(*) per treat outcome category |
| Denominator | COUNT(*) where treat outcome IS NOT NULL |
| Disaggregations | treat outcome (6 categories), book number, tb reg date (monthly), sex, disease site, treatment regimen |
| Visualization | 100% stacked bar chart - X-axis: book number or time period; Y-axis: percentage share; fill: treat outcome. Separate panel for absolute counts. |
| Data Source | mw tb refinery tb reg |

40.4 Lost to follow-up rate

Table 40.4: 4. Lost to follow-up rate

| | |
|------------------------|---|
| Goal | Identify facilities or time periods with elevated rates of patients lost to follow-up, enabling targeted retention interventions. |
| Numerator | COUNT(*) where treat outcome = 'Lost to Follow-up' |
| Denominator | COUNT(*) where treat outcome IS NOT NULL |
| Disaggregations | book number, tb reg date (quarterly), sex, risk group, treatment regimen |
| Visualization | Line chart - X-axis: tb reg date by quarter; Y-axis: percentage lost to follow-up. One line per book number for comparison. |
| Data Source | mw tb refinery tb reg |

40.5 Mortality rate during treatment

Table 40.5: 5. Mortality rate during treatment

| | |
|------------------------|--|
| Goal | Track the proportion of registered TB patients who died during treatment. Supports clinical quality review and facility benchmarking. |
| Numerator | COUNT(*) where treat outcome = 'Dead' |
| Denominator | COUNT(*) where treat outcome IS NOT NULL |
| Disaggregations | book number, tb reg date (quarterly), sex, age years (grouped), disease site, hiv status |
| Visualization | Bar chart - X-axis: book number; Y-axis: mortality rate (%). Colour bars by hiv status split for HIV-attributable mortality exploration. |
| Data Source | mw tb refinery tb reg |

40.6 Moved to second-line treatment

Table 40.6: 6. Moved to second-line treatment

| | |
|------------------|--|
| Goal | Monitor the proportion of patients escalated to second-line treatment, indicating drug resistance or treatment failure burden. |
| Numerator | COUNT(*) where second line = 'Yes' |

| | |
|------------------------|--|
| Denominator | COUNT(*) from mw tb refinery tb reg where tb reg date IS NOT NULL |
| Disaggregations | |
| Visualization | |
| Data Source | book number, tb reg date (quarterly), treatment regimen, dst result Single KPI tile (percentage) with trend sparkline. Secondary: table by book number. |
| Data Source | mw tb refinery tb reg |

40.7 Patient age and sex profile

Table 40.7: 7. Patient age and sex profile

| | |
|----------------------------|---|
| Goal | Characterise the demographic profile of registered TB patients to inform targeted outreach and resource allocation. |
| Definition of Terms | age years is a numeric value. Recommended age groups: 0-4, 5-14, 15-24, 25-34, 35-44, 45-54, 55-64, 65+. |
| Numerator | COUNT(*) per age group and sex |
| Denominator | N/A - absolute counts |
| Disaggregations | sex ('Male', 'Female'), age years (grouped), book number, tb reg date (annual) |
| Visualization | Population pyramid - Y-axis: age group; X-axis: count, mirrored by sex (Male left, Female right). |
| Data Source | mw tb refinery tb reg |

40.8 Occupation and risk group distribution

Table 40.8: 8. Occupation / risk group distribution

| | |
|------------------------|--|
| Goal | Identify which occupational or risk groups contribute the most TB registrations, to prioritise targeted screening and contact tracing (particularly for mining communities). |
| Numerator | COUNT(*) per risk group value |
| Denominator | COUNT(*) where risk group IS NOT NULL |
| Disaggregations | risk group ('Miner', 'Ex-Miner', 'Mining Community', 'HCW', 'Prisoner', 'HH Contact', 'Migrant', 'Other'), book number, tb reg date (annual) |
| Visualization | Horizontal bar chart - Y-axis: risk group; X-axis: count or percentage. Sort bars by descending count. |
| Data Source | mw tb refinery tb reg |

40.9 Previously treated patients

Table 40.9: 9. Previously treated patients

| | |
|----------------------------|--|
| Goal | Monitor the burden of retreatment cases, which carry higher risk of drug resistance and poorer outcomes. |
| Definition of Terms | Previously treated: treatment history IN ('Relap', 'RALF', 'Fail', 'Other'). New case: treatment history = 'New'. |

| | |
|------------------------|--|
| Numerator | COUNT(*) where treatment history IN ('Relap', 'RALF', 'Fail', 'Other') |
| Denominator | COUNT(*) where treatment history IS NOT NULL |
| Disaggregations | treatment history (individual category), book number, tb reg date (quarterly), sex |
| Visualization | Stacked bar chart - X-axis: time period; Y-axis: count; stack: New vs Previously Treated (further split by retreatment reason in a drill-down view). |
| Data Source | mw tb refinery tb reg |

41 Bacteriology & Diagnosis

41.1 Bacteriological confirmation rate

Table 41.1: 10. Bacteriological confirmation rate

| | |
|------------------------|---|
| Goal | Measure the proportion of TB patients diagnosed bacteriologically vs clinically. Higher bacteriological confirmation rates indicate stronger diagnostic capacity. |
| Numerator | COUNT(*) where how diagnosed = 'Bacteriologically' |
| Denominator | COUNT(*) where how diagnosed IS NOT NULL |
| Disaggregations | book number, tb reg date (quarterly), disease site, sex |
| Visualization | Dual KPI tiles (% bacteriological, % clinical) with time-series line chart below showing trend. |
| Data Source | mw tb refinery tb reg |

41.2 Xpert MTB positivity rate

Table 41.2: 11. Xpert MTB positivity rate

| | |
|------------------------|---|
| Goal | Track the proportion of Xpert tests returning MTB Detected to monitor TB burden in tested patients and Xpert utilisation. |
| Numerator | COUNT(*) where xpert result = 'MTB Detected' |
| Denominator | COUNT(*) where xpert result IS NOT NULL |
| Disaggregations | book number, tb reg date (monthly), sex, disease site |
| Visualization | Line chart - X-axis: tb reg date by month; Y-axis: Xpert positivity rate (%); one line per book number. |
| Data Source | mw tb refinery tb reg |

41.3 Drug-resistant TB rate (DST)

Table 41.3: 12. Drug-resistant TB rate (DST)

| | |
|----------------------------|---|
| Goal | Identify the proportion of tested patients with RIF-resistant TB, driving decisions on second-line treatment capacity and drug procurement. |
| Definition of Terms | Tested: dst date IS NOT NULL. RIF resistant: dst result = 'RIF Resistant Detected'. |
| Numerator | COUNT(*) where dst result = 'RIF Resistant Detected' |
| Denominator | COUNT(*) where dst date IS NOT NULL |

| | |
|------------------------|---|
| Disaggregations | book number, tb reg date (quarterly), sex, treatment regimen, treatment history |
| Visualization | Bar chart - X-axis: book number; Y-axis: % RIF resistant. Stack DST result categories in an absolute count panel alongside. |
| Data Source | mw tb refinery tb reg |

41.4 Smear conversion rate at 2 months

Table 41.4: 13. Smear conversion rate at 2 months

| | |
|----------------------------|---|
| Goal | Measure treatment response by tracking the proportion of initially smear-positive patients who convert to negative at 2 months. A key early treatment efficacy signal. |
| Definition of Terms | Smear positive at baseline: smear result not in ('Negative', NULL, 'Invalid'). Converted at 2 months: smear result 2 = 'Negative'. Only patients with both baseline and 2-month smear recorded are included. |
| Numerator | COUNT(*) where smear result IS NOT NULL AND smear result NOT IN ('Negative', 'Invalid') AND smear result 2 = 'Negative' |
| Denominator | COUNT(*) where smear result IS NOT NULL AND smear result NOT IN ('Negative', 'Invalid') AND smear result 2 IS NOT NULL |
| Disaggregations | book number, tb reg date (quarterly), treatment regimen, hiv status |
| Visualization | KPI tile (% converted) with trend line. Secondary: table showing median time to conversion by treatment regimen. |
| Data Source | mw tb refinery tb reg |

42 Treatment Regimen

42.1 Treatment regimen distribution

Table 42.1: 14. Treatment regimen distribution

| | |
|------------------------|--|
| Goal | Monitor the uptake of newer regimens (BPalM, BPal) relative to the standard 2RHZE/4RH regimen. Tracks programmatic transition to shorter regimens. |
| Numerator | COUNT(*) per treatment regimen value |
| Denominator | COUNT(*) where treatment regimen IS NOT NULL |
| Disaggregations | treatment regimen ('2RHZE/4RH', 'BPalM', 'BPal', 'Other Regimen'), book number, tb reg date (quarterly), disease site, dst result |
| Visualization | 100% stacked bar chart - X-axis: tb reg date by quarter; Y-axis: proportion; fill: treatment regimen. Absolute count panel alongside. |
| Data Source | mw tb refinery tb reg |

42.2 DOT and adherence support coverage

Table 42.2: 15. DOT and adherence support coverage

| | |
|------------------------|--|
| Goal | Measure the proportion of patients receiving documented DOT/adherence support and identify the most common support type. Supports community health worker workload planning. |
| Numerator | COUNT(*) where adherence support IS NOT NULL |
| Denominator | COUNT(*) from mw tb refinery tb reg where tb reg date IS NOT NULL |
| Disaggregations | adherence support ('Guardian', 'Health Surveillance Assistant', 'Volunteer', 'Healthcare Worker'), book number, tb reg date (quarterly) |
| Visualization | Donut chart for support type split. Secondary: line chart of coverage rate over time. |
| Data Source | mw tb refinery tb reg |

43 HIV & ART

43.1 HIV co-infection rate among TB patients

Table 43.1: 16. HIV co-infection rate among TB patients

| | |
|------------------------|--|
| Goal | Track the burden of HIV among registered TB patients. High TB/HIV co-infection drives the need for integrated HIV services at TB facilities. |
| Numerator | COUNT(*) where hiv status = 'Positive' |
| Denominator | COUNT(*) where hiv status IS NOT NULL |
| Disaggregations | hiv status ('Positive', 'Negative', 'Unknown'), book number, tb reg date (quarterly), sex, age years (grouped), disease site |
| Visualization | Stacked bar chart - X-axis: time period; Y-axis: count; stack: Positive / Negative / Unknown. Add percentage label on Positive segment. |
| Data Source | mw tb refinery tb reg |

43.2 HIV testing coverage among TB patients

Table 43.2: 17. HIV testing coverage among TB patients

| | |
|------------------------|--|
| Goal | Measure the proportion of TB patients who received an HIV test result (not Unknown), assessing the completeness of integrated HIV testing. |
| Numerator | COUNT(*) where hiv status IN ('Positive', 'Negative') |
| Denominator | COUNT(*) where hiv status IS NOT NULL |
| Disaggregations | book number, tb reg date (quarterly), hiv test time ('Before TB Reg', 'After TB Reg'), sex |
| Visualization | KPI tile (% tested, % positive) with trend line. Secondary: bar chart split by hiv test time. |
| Data Source | mw tb refinery tb reg |

43.3 ART initiation among HIV-positive TB patients

Table 43.3: 18. ART initiation among HIV-positive TB patients

| | |
|----------------------------|--|
| Goal | Measure the proportion of HIV-positive TB patients who started ART, and the timing of ART relative to TB treatment. ART during TB treatment reduces mortality. |
| Definition of Terms | ART initiated: arv status IN ('Before TB Tx', 'While on TB Tx'). Not initiated: arv status = 'Not While on TB Tx' or arv status IS NULL (among HIV-positive). |

| | |
|------------------------|---|
| Numerator | COUNT(*) where hiv status = 'Positive' AND arv status IN ('Before TB Tx', 'While on TB Tx') |
| Denominator | COUNT(*) where hiv status = 'Positive' |
| Disaggregations | arv status (3 categories), book number, tb reg date (quarterly), sex |
| Visualization | Stacked bar chart - X-axis: time period; stack: Before TB Tx / While on TB Tx / Not While on TB Tx. Add 90% WHO target line for ART coverage. |
| Data Source | mw tb refinery tb reg |

43.4 CPT coverage among HIV-positive TB patients

Table 43.4: 19. CPT coverage among HIV-positive TB patients

| | |
|------------------------|---|
| Goal | Monitor cotrimoxazole preventive therapy (CPT) uptake among HIV-positive TB patients, reducing mortality from opportunistic infections. |
| Numerator | COUNT(*) where hiv status = 'Positive' AND cpt status = 'Yes' |
| Denominator | COUNT(*) where hiv status = 'Positive' |
| Disaggregations | book number, tb reg date (quarterly), sex, treat outcome |
| Visualization | KPI tile (CPT coverage %). Secondary: line chart of coverage rate over time per book number. |
| Data Source | mw tb refinery tb reg |

44 Post-TB / PRP

44.1 PRP enrolment rate among eligible patients

Table 44.1: 20. PRP enrolment rate among eligible patients

| | |
|----------------------------|---|
| Goal | Track the proportion of cured or completed patients enrolled into the Post-TB Lung Disease (PRP) programme, monitoring uptake of post-treatment care. |
| Definition of Terms | Eligible: treat outcome IN ('Cured', 'Completed'). Enrolled: enrolled prp = 'Yes'. |
| Numerator | COUNT(*) where treat outcome IN ('Cured', 'Completed') AND enrolled prp = 'Yes' |
| Denominator | COUNT(*) where treat outcome IN ('Cured', 'Completed') |
| Disaggregations | book number, tb reg date (quarterly), sex, disease site, treatment regimen |
| Visualization | KPI tile (% enrolled) with trend sparkline. Secondary: bar chart by book number to compare facility enrolment rates. |
| Data Source | mw tb refinery tb reg |

44.2 PTLTD diagnosis method distribution

Table 44.2: 21. PTLTD diagnosis method distribution

| | |
|------------------------|--|
| Goal | Understand which diagnostic methods are being used to identify Post-TB Lung Disease among PRP-enrolled patients, supporting clinical protocol adherence. |
| Numerator | COUNT(*) per ptld how diagnosed value where enrolled prp = 'Yes' |
| Denominator | COUNT(*) where enrolled prp = 'Yes' AND ptld how diagnosed IS NOT NULL |
| Disaggregations | ptld how diagnosed ('Spirometry/LFT', 'Persistent Respiratory Symptoms', 'Abnormalities X-ray', 'Reduced Exercise'), book number |
| Visualization | Horizontal bar chart sorted by frequency. One panel per book number for facility comparison. |
| Data Source | mw tb refinery tb reg |

44.3 PRP programme outcome distribution

Table 44.3: 22. PRP programme outcome distribution

| | |
|------------------------|---|
| Goal | Monitor outcomes for patients enrolled in the PRP programme to assess programme effectiveness and identify patients still active or lost. |
| Numerator | COUNT(*) per prp outcome value |
| Denominator | COUNT(*) where prp outcome IS NOT NULL |
| Disaggregations | prp outcome ('Completed', 'Died', 'Lost to Follow-up', 'Active TB', 'Excluded'), book number, tb reg date (annual) |
| Visualization | Stacked bar chart or donut chart per facility. Flag 'Active TB' recurrence as a clinical alert metric. |
| Data Source | mw tb refinery tb reg |

45 Data Quality

45.1 Check failure rate by book

Table 45.1: 23. Check failure rate by book (facility)

| | |
|------------------------|---|
| Goal | Identify register books (facilities) with the highest rate of data quality check failures, guiding targeted data quality improvement support. |
| Numerator | COUNT(DISTINCT record identifier) where severity IN ('Error', 'Warning') from mw tb refinery tb reg checks |
| Denominator | COUNT(DISTINCT Record number || Page ID) from mw tb refinery tb reg |
| Disaggregations | book number, severity ('Error', 'Warning'), tags (check category), Submission Date (monthly) |
| Visualization | Bar chart - X-axis: book number; Y-axis: check failure rate (%). Stack Error and Warning in different colours. Sort by descending Error rate. |
| Data Source | mw tb refinery tb reg checks joined to mw tb refinery tb reg on Page ID + Record number |

45.2 Most frequent check failures

Table 45.2: 24. Most frequent check failures

| | |
|------------------------|---|
| Goal | Rank data quality issues by frequency to identify systemic problems (e.g. a field consistently left blank or a question frequently multi-selected) requiring training or form revision. |
| Numerator | COUNT(*) per check message from mw tb refinery tb reg checks |
| Denominator | N/A - absolute counts ranked |
| Disaggregations | check message (specific failure description), severity, tags, book number, Submission Date (monthly) |
| Visualization | Ranked horizontal bar chart - top 10 most frequent check message values. Colour by severity (Error = red, Warning = amber). |
| Data Source | mw tb refinery tb reg checks |

45.3 Records excluded due to Errors

Table 45.3: 25. Records excluded due to Errors

| | |
|--|--|
| | |
|--|--|

| | |
|------------------------|--|
| Goal | Monitor the proportion of submitted records excluded from analysis due to Error-severity check failures, as a summary measure of data entry quality over time. |
| Numerator | COUNT(DISTINCT Page ID || Record number) from mw tb refinery tb reg checks where severity = 'Error' |
| Denominator | COUNT(DISTINCT Page ID || Record number) from mw tb refinery tb reg UNION COUNT from mw tb refinery tb reg checks where severity = 'Error' |
| Disaggregations | book number, Submission Date (monthly), tags (error type: invalid-date-check vs clinical-logic-check) |
| Visualization | Line chart - X-axis: Submission Date by month; Y-axis: % records excluded. One line per book number. Alert threshold line at 5%. |
| Data Source | mw tb refinery tb reg checks |

46 Indicator summary table

Table 46.1: All 25 dashboard indicators across 6 thematic groups

| # | Indicator Name | Group | Type | Refinery Model |
|----|--------------------------------------|--------------------------|----------|------------------------------|
| 1 | Total patients registered | Programme Performance | Count | mw tb refinery tb reg |
| 2 | Treatment success rate | Programme Performance | Rate / % | mw tb refinery tb reg |
| 3 | Treatment outcome distribution | Programme Performance | Rate / % | mw tb refinery tb reg |
| 4 | Lost to follow-up rate | Programme Performance | Rate / % | mw tb refinery tb reg |
| 5 | Mortality rate during treatment | Programme Performance | Rate / % | mw tb refinery tb reg |
| 6 | Moved to second-line treatment | Programme Performance | Rate / % | mw tb refinery tb reg |
| 7 | Patient age and sex profile | Programme Performance | Count | mw tb refinery tb reg |
| 8 | Occupation / risk group distribution | Programme Performance | Rate / % | mw tb refinery tb reg |
| 9 | Previously treated patients | Programme Performance | Rate / % | mw tb refinery tb reg |
| 10 | Bacteriological confirmation rate | Bacteriology & Diagnosis | Rate / % | mw tb refinery tb reg |
| 11 | Xpert MTB positivity rate | Bacteriology & Diagnosis | Rate / % | mw tb refinery tb reg |
| 12 | Drug-resistant TB rate (DST) | Bacteriology & Diagnosis | Rate / % | mw tb refinery tb reg |
| 13 | Smear conversion rate at 2 months | Bacteriology & Diagnosis | Rate / % | mw tb refinery tb reg |
| 14 | Treatment regimen distribution | Treatment Regimen | Rate / % | mw tb refinery tb reg |
| 15 | DOT and adherence support coverage | Treatment Regimen | Rate / % | mw tb refinery tb reg |
| 16 | HIV co-infection rate | HIV & ART | Rate / % | mw tb refinery tb reg |
| 17 | HIV testing coverage | HIV & ART | Rate / % | mw tb refinery tb reg |
| 18 | ART initiation among HIV-positive | HIV & ART | Rate / % | mw tb refinery tb reg |
| 19 | CPT coverage among HIV-positive | HIV & ART | Rate / % | mw tb refinery tb reg |
| 20 | PRP enrolment rate | Post-TB / PRP | Rate / % | mw tb refinery tb reg |
| 21 | PTLD diagnosis method distribution | Post-TB / PRP | Rate / % | mw tb refinery tb reg |
| 22 | PRP programme outcome distribution | Post-TB / PRP | Rate / % | mw tb refinery tb reg |
| 23 | Check failure rate by book | Data Quality | Rate / % | mw tb refinery tb reg checks |
| 24 | Most frequent check failures | Data Quality | Count | mw tb refinery tb reg checks |
| 25 | Records excluded due to Errors | Data Quality | Rate / % | mw tb refinery tb reg checks |