



PRACTICAL QUALITY TOOL

Data Quality, Cleaning and Handover Checklist

A practical checklist for research, evaluation and monitoring teams working with survey, qualitative or mixed-method datasets.

Use this checklist when you need to:

- audit an existing survey or monitoring dataset
- document cleaning decisions
- verify derived indicators and disaggregations
- prepare reproducible analysis outputs
- organise a secure, usable handover package

Gesem Research and Consulting PLC

Researching Today for a Better Future

PHASE 1

Data intake and validation

Before cleaning starts, establish what the data represent, which version is authoritative and how the expected indicators should be constructed. The aim is to prevent undocumented assumptions from entering the analysis.

- [] Confirm the authoritative source files**
Record file names, versions, dates received, formats and the person responsible for confirming the analysis copy.
- [] Preserve an untouched raw-data copy**
Keep a read-only source file so every subsequent transformation can be traced back to the original data received.
- [] Review variable names, labels and coding**
Check whether names, value labels, missing-value codes and response categories are consistent with the questionnaire and codebook.
- [] Validate respondent and record identifiers**
Check uniqueness, duplicate identifiers, household/respondent relationships and merge keys before combining datasets.
- [] Check completeness and missingness**
Review item non-response, system missing values and patterns by geography, interviewer, date or respondent group.
- [] Test skip patterns and internal logic**
Identify impossible combinations, skipped mandatory items, contradictory answers and records that violate questionnaire logic.
- [] Review ranges and outliers**
Separate true extreme values from entry, coding or unit errors. Document any capping, correction or exclusion rule.
- [] Reconcile sample and fieldwork records**
Where available, compare final data against sample lists, call sheets, submission logs or supervisor records to identify unexplained gaps.

Decision point

Do not begin indicator production until the team agrees which records are valid, which corrections are permissible and which issues must remain transparently flagged.

Cleaning, analysis and documentation

Cleaning should be reproducible, indicator calculations should be independently checkable and every material analytical decision should be documented.

[] Write cleaning rules before applying them

State what constitutes an error, how corrections will be made and when records will be retained, flagged or excluded.

[] Use reproducible syntax or scripts

Where possible, avoid irreversible manual edits. Keep syntax, formulas or scripts that can recreate the cleaned dataset from the raw copy.

[] Maintain a cleaning and decision log

Record corrections, exclusions, recodes, derived variables and reasons for non-routine decisions.

[] Verify derived indicators

Recalculate key indicators independently or with spot checks. Confirm denominators, weights, thresholds, time periods and aggregation rules.

[] Check disaggregations and small cells

Review sex, age, disability, geography and other planned disaggregations for coding consistency and disclosure or interpretation risks.

[] Compare distributions before and after cleaning

Look for unexpected shifts caused by recoding, exclusions or transformations, especially for primary outcomes.

[] Align analysis with the evaluation or research questions

Tables and models should answer defined questions rather than simply report every available variable.

[] Integrate qualitative evidence transparently

Use coding frameworks, theme matrices or structured synthesis so narrative interpretation can be traced to source material.

[] Separate findings from interpretation

Keep factual results, analytical explanation and recommendations conceptually distinct so clients can review the evidence chain.

Minimum documentation	Useful additions
Clean dataset Cleaning log Codebook/data dictionary Indicator definitions Analysis syntax or formulas	Quality-control memo Version register Derived-variable specification Qualitative code matrix Table/figure index

Handover and acceptance

A clean dataset is not a complete handover. The receiving team should be able to understand what was done, reproduce critical outputs and know which limitations remain.

- [] Freeze and label the final analytical dataset**
Use an unambiguous final file name, version number and date. Keep working files separate from the accepted handover package.
- [] Provide a usable codebook or data dictionary**
Define variable names, labels, values, units, missing codes, derived fields and any special interpretation notes.
- [] Package analysis files with dependencies**
Include syntax, scripts, lookup tables and supporting files needed to reproduce key tables, indicators or models.
- [] Document unresolved quality limitations**
State remaining missingness, sampling limitations, measurement concerns or records that could not be verified.
- [] Check privacy and disclosure risk before transfer**
Remove unnecessary direct identifiers, confirm access permissions and use an agreed secure transfer method.
- [] Match tables and figures to the final data version**
Ensure report tables, dashboards and presentations were generated from the same accepted analytical dataset.
- [] Confirm client acceptance criteria**
Agree what constitutes complete handover: files, formats, documentation, revisions, storage arrangements and ownership.
- [] Retain only what is authorised**
After acceptance, follow the agreed retention, archiving or deletion protocol for raw and working data.

Recommended final handover package

- 01.** Clean analytical dataset **02.** Codebook/data dictionary **03.** Cleaning and decision log
04. Syntax/scripts **05.** Indicator tables **06.** Quality/limitations note **07.** Agreed visual outputs

Need support with data cleaning, analysis or handover?

Gesem supports research and evaluation teams with data audits, cleaning, indicator reconstruction, quantitative and qualitative analysis, documentation and decision-ready reporting.

Website: <https://grcplc.com/service-data-management-analytics.php>

Email: info@grcplc.com **Phone:** +251 111 54 54 44