PORTFOLIO PROJECT

HIV Treatment, Clinic Visit and Viral Load Data Quality Assessment (DQA)

A Practical Data Quality Assessment Using Microsoft Excel and PowerPoint

Author: Confidence Ogochukwu Anyaorah, MPH

Tools: Microsoft Excel and Microsoft PowerPoint

Illustrative Project Period: July–September 2026

Data Source: Synthetic data created for portfolio demonstration

Important Note on Data:

I used entirely synthetic data, including invented facility names, figures, dates, and findings. No patient records or real facility names are included.

Executive Summary

Reliable HIV programme data help managers make informed decisions about commodity planning, staffing, patient follow-up, and service delivery. Inaccurate, incomplete, or delayed reporting can undermine these decisions and affect programme performance.

Project Overview: This practice Data Quality Assessment (DQA) used a synthetic dataset of 180 records to assess data concordance, reporting timeliness, and discrepancies between reported figures and source registers.

Key Findings: Of the 180 records reviewed, 160 matched the source registers, while 20 were flagged for discrepancies: 4 over-reported records, 13 under-reported records, and 3 missing records. Overall concordance was 98.1%, and 88.1% of values were reported on time.

The findings highlighted several potential data quality issues, including the absence of pre-submission verification, delayed register updates, unrecorded refill visits, calculation errors, and faulty laptop equipment. Unrecorded refill visits were a particularly important concern because they could result in active clients being incorrectly classified as interrupted in treatment (IIT), potentially affecting patient follow-up and antiretroviral medicine planning.

Recommended Actions: The assessment proposes 14 practical corrective actions to address the identified gaps, strengthen routine data verification, improve reporting timeliness, and support more reliable HIV programme data.

Project at a Glance

Item	Detail
Programme area	HIV treatment, clinic visits (drug refills) and viral load reporting
Scope	10 made-up facilities x 3 months x 6 indicators = 180 records
Quality checks	Completeness, timeliness, accuracy (concordance), consistency
Method	Verify, compare, classify, find causes, recommend, monitor
Main result	160 of 180 records matched; 20 flagged; clinic visits and IIT risk highlighted
Tools	Excel for cleaning, formulas, summaries and charts; PowerPoint for findings

1. Introduction, Objectives and Problems

Every month, health facilities report how many people are on HIV treatment, how many came to the clinic for a drug refill, and how many had a viral load test. Programmes use these numbers to decide where to send drugs, staff and support. If the numbers are wrong, late or missing, decisions can be wrong too.

A Data Quality Assessment (DQA) is a simple check. We compare what facilities report with what is written in their registers, and we check that reports arrive on time and are complete. This project shows how I carry out a DQA using Excel, and how I share the findings in PowerPoint. I wrote it in plain language so that anyone can follow it.

My objectives

1. **Check accuracy.** Find out whether the numbers facilities report match their registers.
2. **Check completeness and timeliness.** Find out how many numbers were sent, and how many arrived on time.
3. **Find where the problems are.** Show which facilities, indicators and months need attention.
4. **Find the causes.** Understand why numbers do not match or arrive late.
5. **Recommend practical actions.** Suggest clear steps that facility and programme teams can follow.
6. **Summarise it for managers.** Build an Excel dashboard and a short PowerPoint summary.

Problems I want to solve

Problem	Why it matters
Reported numbers differ from the register	The programme may think more or fewer people are on treatment than there really are.
Clinic (refill) visits are under- or over-reported	Drugs are planned from the wrong numbers, and active clients can wrongly look lost to follow-up (IIT).
Reports are late	Decisions are made without the latest information.
Reports or values are missing	The picture is incomplete, and comparisons between facilities are unfair.
Viral load numbers do not fit together	For example, more results than samples collected. We cannot trust the viral load picture.
The same patient is counted twice or not at all	Counts are inflated or too low.

2. Scope and Data Used

This practice assessment covered 10 fictional health facilities, identified as SYN001 to SYN010, over three reporting months: July, August, and September 2026.

Six indicators were assessed at each facility for each month, producing a total of 180 facility-month-indicator records (10 facilities × 3 months × 6 indicators).

Indicators Assessed

Indicator (code)	What it counts
Patients on treatment (TX_CURR)	The number of clients currently receiving HIV treatment at the reporting point.
New patients started (TX_NEW)	The number of clients who initiated HIV treatment during the reporting period.
Clinic visits (CLINIC_VISITS)	The number of clinic visits at which clients received antiretroviral medicine refills.
Viral load samples (VL_SAMPLES)	The number of viral load samples collected during the reporting period.
Viral load results (VL_RESULTS)	The number of viral load test results received during the reporting period.
Viral load suppressed (VL_SUPPRESSED)	The number of clients whose viral load results are below 1,000 copies/mL.

For every record I kept three numbers:

- **Verified value:** the number confirmed from the facility register. This is the reference.
- **Reported value:** the number the facility sent in its monthly report. Blank means nothing was reported. A blank is not the same as zero.
- **System value:** the number in the facility's electronic record.

The difference between reported and verified is the **variance** (reported minus verified). A positive variance means over-reported. A negative variance means under-reported.

Why clinic visits are included

A clinic visit here means a client who came and was given a drug refill. Programmes watch this number closely because it affects two things: how many drugs are ordered, and which clients are seen as having stopped treatment. If 20 clients came for a refill but only 15 were reported, five clients are missing from the numbers. Section 7 explains what that can cause.

3. How I Assessed Data Quality

I assessed four dimensions of data quality. Each answers a specific question and uses a defined method of measurement.

What I checked	The question it answers	How I checked it
Completeness	Was every expected number sent?	Values received compared with values expected.
Timeliness	Did the report arrive on time?	Date sent compared with the deadline (the 15th of the next month).
Accuracy (concordance)	Did reported figures match the source registers?	Compared each reported value with its corresponding verified value and measured the size of any discrepancy.

What I checked	The question it answers	How I checked it
Consistency	Do the numbers make sense together?	For example, results cannot be more than samples collected.

Each record was given one of these results:

- **Concordant:** the reported number is exactly the same as the verified number.
- **Within tolerance:** a small difference of 5% or less. I counted this as a match, but I still looked at patterns.
- **Over-reported:** more than 5% higher than the register.
- **Under-reported:** more than 5% lower than the register.
- **Missing report:** no number was reported.

The 5% tolerance is an assumption I chose for this exercise. In Excel it sits in one cell, so it can be changed. The calculations are simple:

- **Variance** = reported value minus verified value
- **Concordance rate** = (1 minus the size of the variance divided by the verified value) x 100
- **Match rate** = records that matched, divided by records checked, x 100
- **Completeness** = values received, divided by values expected, x 100
- **Timeliness** = values on time, divided by values received, x 100

Three rules keep the results fair. If the verified value is zero, the percentage is not calculated and the record is marked for review. A missing report counts against completeness but is not given a concordance score. And completeness, timeliness and accuracy are never mixed into one score, because a report can arrive on time and still be wrong.

4. Checking the Data Before Analysis

Before analysing the dataset, I performed basic data validation checks to identify errors, omissions and inconsistencies that could affect the results.

The checks covered:

Missing values: Identified blank fields and kept them separate from genuine zero values. Missing reports were counted against completeness.

Invalid values: Checked for negative numbers and other impossible or unexpected values.

Duplicate records: Checked whether the same facility, reporting month and indicator had been entered more than once.

Date and timeliness errors: Checked date formats and compared submission dates with the assumed reporting deadline.

Logical inconsistencies: Flagged records where suppressed viral load results exceeded total viral load results, or where viral load results exceeded samples collected in the same reporting period. These cases require investigation because results received in a month may relate to samples collected earlier.

Unusual clinic visit figures: Flagged clinic visit counts that appeared unusually high or low relative to the number of clients on treatment for further verification.

Any unusual value was flagged for review rather than automatically corrected. Changes were made only after checking the relevant source records. This helped preserve the distinction between genuine data errors, reporting omissions and differences that required further investigation.

5. The Excel Workbook

I conducted the analysis in a single Excel workbook designed to make the assessment transparent, easy to review, and reproducible. The workbook contains the synthetic dataset, data definitions, configurable assessment rules, validation checks, summary tables, findings, and dashboard.

To distinguish inputs from calculations, manually entered values are formatted in blue text, while formula-driven calculations are formatted in black. The results presented in this portfolio are drawn from the workbook.

Sheet	What it holds
DQA_Dataset	The 180 records, with formulas for variance, concordance, timeliness and result.
Data_Dictionary	A plain explanation of every column and the rules used.
Settings	The 5% tolerance and the deadline day, which can be changed.
Consistency_Checks	Checks that the viral load numbers make sense together.
DQA_Summary	Totals by indicator, facility and month, plus the clinic visit table.
DQA_Findings	Each finding, its cause, the action, who is responsible and the status.
Dashboard	Headline measures and four charts.

Facility_ID	Reporting_Month	Indicator_Name	Reported_Value	System_Value	Verified_Value	Variance	Variance_Pct	Timely	Result
SYN001	2026-07	Patients on treatment	1,250	1,250	1,250	0	0.0%	Yes	Concordant
SYN001	2026-07	New patients started	21	21	21	0	0.0%	Yes	Concordant
SYN001	2026-07	Clinic visits (drug refills)	769	769	769	0	0.0%	Yes	Concordant
SYN001	2026-07	Viral load samples collected	326	326	326	0	0.0%	Yes	Concordant
SYN001	2026-07	Viral load results received	317	317	317	0	0.0%	Yes	Concordant
SYN001	2026-07	Viral load suppressed	296	296	296	0	0.0%	Yes	Concordant
SYN001	2026-08	Patients on treatment	1,265	1,265	1,265	0	0.0%	Yes	Concordant
SYN001	2026-08	New patients started	21	21	21	0	0.0%	Yes	Concordant
SYN001	2026-08	Clinic visits (drug refills)	848	778	778	70	9.0%	Yes	Over-reported
SYN001	2026-08	Viral load samples collected	330	330	330	0	0.0%	Yes	Concordant
SYN001	2026-08	Viral load results received	321	321	321	0	0.0%	Yes	Concordant
SYN001	2026-08	Viral load suppressed	295	300	300	-5	-1.7%	Yes	Within tolerance
SYN001	2026-09	Patients on treatment	1,280	1,280	1,280	0	0.0%	Yes	Concordant
SYN001	2026-09	New patients started	21	21	21	0	0.0%	Yes	Concordant

Figure 1. The DQA_Dataset sheet (first 14 of 180 records). Colours show the result of each check. Synthetic data.

Field	Meaning (plain words)	Allowed values / rule
Record_ID	Row number, 1 to 180.	Unique
Facility_ID	Made-up facility code.	SYN001 to SYN010
Reporting_Month	The month the numbers belong to.	2026-07, 2026-08, 2026-09
Indicator_Code	Short code for what is counted.	TX_CURR, TX_NEW, CLINIC_VISITS, VL_SAMPLES, VL_RESULTS, VL_SUPPRESSED (drop-down list)
Indicator_Name	Plain name of the indicator.	See list on the README and in the project report
Reported_Value	The number the facility sent in its monthly report.	Blank means the report or value is missing (not the same as 0)
System_Value	The number in the facility's electronic record.	Whole number, 0 or more
Verified_Value	The number confirmed from the facility register. This is the reference.	Whole number, 0 or more
Submission_Date	The date the report was sent.	Blank if nothing was submitted
Submission_Deadline	The date the report was due.	Deadline day on Settings, in the month after the reporting month
Variance	How far the report is from the verified value.	Reported minus Verified. Positive = over-reported. Negative = under-reported

Figure 2. The Data_Dictionary sheet (first rows). Synthetic workbook.

6. Findings

The assessment showed high reported numerical concordance across several indicators, alongside specific reporting gaps requiring corrective action. Of the 180 expected facility-month-indicator records, 177 contained reported values and three were missing. Among the 177 reported values, 160 met the predefined matching criterion, while 17 had discrepancies: four were over-reported and 13 were under-reported.

Overall completeness was **98.3%** (177 of 180 expected values). Timeliness was **88.1%** (156 of 177 reported values submitted by the assumed deadline).

These results demonstrate why data quality should be assessed across separate dimensions. High numerical concordance does not eliminate the need to investigate missing values, late submissions, recurring discrepancies and inconsistencies between related indicators.

The findings below examine the results by indicator and facility, followed by an assessment of clinic refill reporting, potential causes and recommended corrective actions. All findings are illustrative and based on synthetic data created for this portfolio project.

Measure	Result
xpected records	180
Reported values received	177 of 180 (98.3% completeness)
Missing values	3
Records meeting the matching criterion	160 of 177 reported values (90.4%); 160 of 180 expected records (88.9%)
Reported records with discrepancies	17: four over-reported and 13 under-reported
Timeliness	156 of 177 reported values submitted on time (88.1%); 21 were late

Interpreting match rate and concordance: Match rate measures the proportion of records that meet a defined matching criterion. Concordance can describe the degree of agreement between reported and verified values, but its percentage must be supported by a clearly stated calculation. The overall concordance figure of 98.1% is therefore retained for verification against the workbook formula before publication.

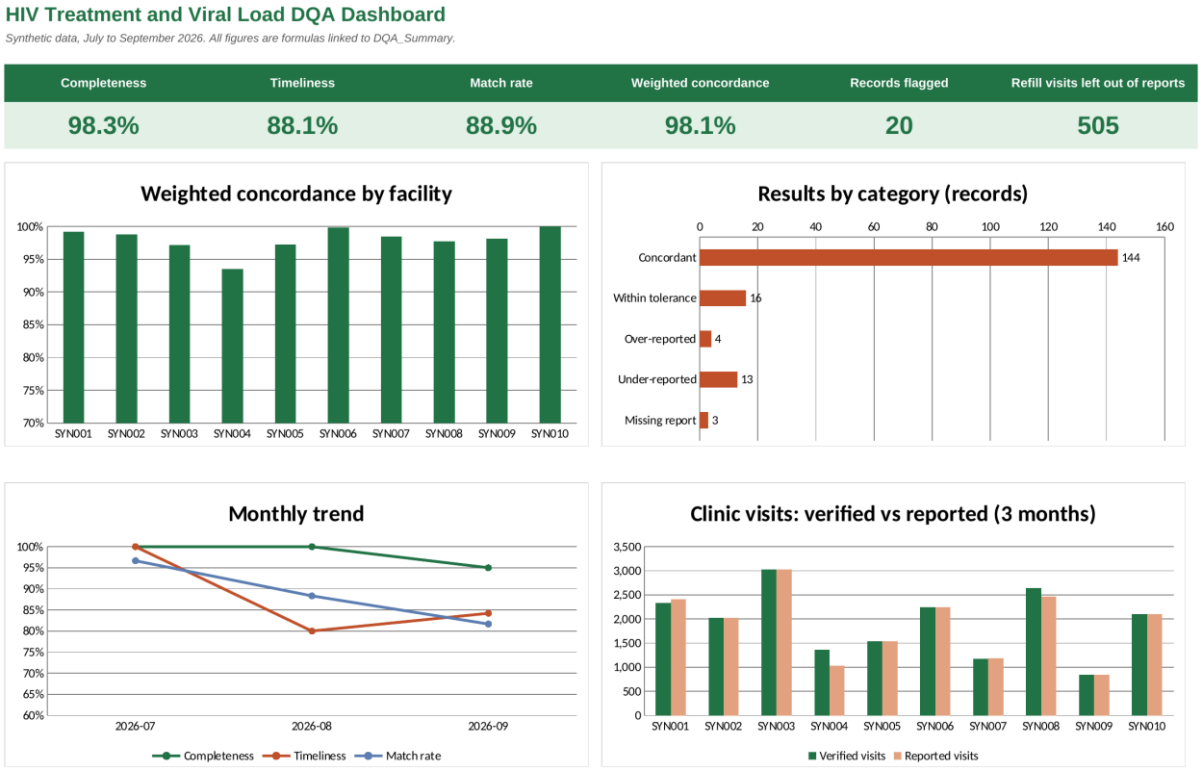


Figure 3. The Excel dashboard: headline measures and four charts (synthetic data).

6.1 By indicator

Indicator	Reported	Verified	Net variance	Concordance	Flagged
Patients on treatment	29,972	30,299	-327	98.2%	5
New patients started	675	658	+17	97.4%	2
Clinic visits (drug refills)	18,863	19,290	-427	97.0%	5
Viral load samples collected	7,821	7,898	-77	99.0%	3
Viral load results received	7,486	7,543	-57	98.2%	4
Viral load suppressed	7,151	7,165	-14	99.7%	1

Note: Reported and verified totals compare only records for which a numerical value was reported. Missing values are excluded from these totals. Net variance is calculated as reported minus verified; a negative value indicates under-reporting, while a positive value indicates over-reporting.

The largest net under-reporting by volume occurred in clinic visits, with 427 fewer visits reported than verified, followed by patients on treatment, with a net difference of 327. These differences warrant investigation because inaccurate reporting can affect programme performance interpretation and operational planning.

Viral load indicators showed relatively high reported concordance, but the flagged records and logical inconsistencies still require review. Aggregate figures alone cannot establish whether the differences arose from recording, transcription, reporting-period alignment or other causes.

Indicator	Reported total	Verified total (same rows)	Net variance	Weighted concordance	Over-reported	Under-reported	Missing	Flagged
Patients on treatment	29,972	30,299	-327	98.2%	0	5	0	5
New patients started	675	658	17	97.4%	2	0	0	2
Clinic visits (drug refills)	18,863	19,290	-427	97.0%	1	4	0	5
Viral load samples collected	7,821	7,898	-77	99.0%	0	2	1	3
Viral load results received	7,486	7,543	-57	98.2%	1	2	1	4
Viral load suppressed	7,151	7,165	-14	99.7%	0	0	1	1
All indicators	**71,968**	**72,853**	**-885**	**98.1%**	**4**	**13**	**3**	**20**

Figure 4. Summary by indicator, from the DQA_Summary sheet (synthetic data).

6.2 By facility

Facility	Flagged records	Concordance	Timeliness	Status
SYN001	1	99.2%	100.0%	Watch
SYN002	0	98.8%	100.0%	On track
SYN003	3	97.1%	100.0%	Needs support
SYN004	3	93.5%	100.0%	Needs support
SYN005	4	97.2%	100.0%	Needs support
SYN006	2	99.8%	33.3%	Watch
SYN007	0	98.5%	100.0%	On track
SYN008	2	97.7%	100.0%	Watch
SYN009	5	98.2%	40.0%	Needs support
SYN010	0	100.0%	100.0%	On track

acility status rules: A facility is classified as Needs support when it has three or more flagged records. Watch applies when it has one or two flagged records, or any late value, provided it does not meet the Needs support threshold. On track applies when there are no flagged records and no late values.

Four facilities were classified as needing support, three were placed on watch, and three were on track. SYN009 had the highest number of flagged records (five) and low timeliness (40.0%). SYN006 also had a timeliness concern, with only 33.3% of reported values submitted on time, despite its high concordance figure.

These results illustrate why numerical agreement alone is insufficient for assessing reporting performance. Facilities with recurring discrepancies or late submissions may need targeted mentoring, improved reporting checks and follow-up on corrective actions.

B. By facility									
Facility	Records	Over-reported	Under-reported	Missing	Flagged	Match rate	Weighted concordance	Timeliness	Status
SYN001	18	1	0	0	1	94.4%	99.2%	100.0%	Watch
SYN002	18	0	0	0	0	100.0%	98.8%	100.0%	On track
SYN003	18	0	3	0	3	83.3%	97.1%	100.0%	Needs support
SYN004	18	0	3	0	3	83.3%	93.5%	100.0%	Needs support
SYN005	18	0	4	0	4	77.8%	97.2%	100.0%	Needs support
SYN006	18	2	0	0	2	88.9%	99.8%	33.3%	Watch
SYN007	18	0	0	0	0	100.0%	98.5%	100.0%	On track
SYN008	18	1	1	0	2	88.9%	97.7%	100.0%	Watch
SYN009	18	0	2	3	5	72.2%	98.2%	40.0%	Needs support
SYN010	18	0	0	0	0	100.0%	100.0%	100.0%	On track
All	**180**	**4**	**13**	**3**	**20**	**88.9%**	**98.1%**	**88.1%**	
Status rule: Needs support = 3 or more flagged records. Watch = 1 or 2 flagged records, or any late value. Otherwise On track.									

Figure 5. Summary by facility, from the DQA_Summary sheet (synthetic data).

6.3 Key observations

- SYN003 – Persistent under-reporting of patients on treatment: Patients on treatment were under-reported by approximately 7% in each of the three months. A repeated pattern may indicate a consistent gap in the counting, reconciliation or reporting process rather than an isolated error. Source records should be reviewed to establish the cause.
- SYN005 – Viral load reporting inconsistencies: Viral load samples and results were under-reported in August and September, while the suppressed count remained unchanged. This resulted in suppressed results exceeding total results received, indicating a logical inconsistency that requires verification.
- SYN006 – Possible over-reporting and late submission: New patients started were over-reported by approximately 25% in August and September. Duplicate entries are one possible explanation, but this must be confirmed by checking source records. Both reports were also submitted late.
- SYN008 – Viral load figures requiring reconciliation: More viral load results were reported than samples collected in September. This should be investigated, taking into account that results received during a reporting month may relate to samples collected in an earlier month.
- SYN009 – Missing values, under-reporting and delays: Three viral load values were missing in September. Patients on treatment were also under-reported in August and September, and submissions were late in both months. The missing values and reporting differences require follow-up to determine their causes.
- SYN002 and SYN007 – Recurring differences within tolerance: Both facilities showed small differences of approximately 3% in a consistent direction across the three months. Although the differences remained within the assumed 5% tolerance, the repeated pattern warrants review. Trend analysis can identify systematic differences that a threshold-based check may not flag.
- These observations are illustrative findings from the synthetic dataset. The patterns indicate where further verification would be appropriate; they do not independently establish the underlying causes.

6.4 Do the numbers fit together?

- **SYN005:** The reported number of suppressed results exceeded the number of viral load results received in August and September, producing implied suppression rates of 120% and 118%, respectively. These figures are logically inconsistent if the counts refer to the same cohort and reporting period.
- **SYN008:** Viral load results exceeded samples collected in September. The figures should be reconciled against laboratory records and reporting-period definitions before concluding that an error occurred.
- **SYN009:** Missing viral load values in September prevented the relevant checks from being completed. The missing values need to be resolved before the affected indicators can be fully assessed.

These checks help identify figures that may be incomplete or inconsistent before they are used for programme decisions. A flagged result is a prompt for investigation, not automatic proof of an error.

Facility_ID	Month	Check 1: results vs samples	Check 2: suppressed vs results	Implied suppression	Overall
SYN001	2026-07	OK	OK	93.4%	OK
SYN001	2026-08	OK	OK	91.9%	OK
SYN001	2026-09	OK	OK	93.5%	OK
SYN002	2026-07	OK	OK	94.6%	OK
SYN002	2026-08	OK	OK	94.2%	OK
SYN002	2026-09	OK	OK	94.3%	OK
SYN003	2026-07	OK	OK	95.2%	OK
SYN003	2026-08	OK	OK	95.0%	OK
SYN003	2026-09	OK	OK	95.0%	OK
SYN004	2026-07	OK	OK	96.2%	OK
SYN004	2026-08	OK	OK	96.3%	OK
SYN004	2026-09	OK	OK	96.3%	OK
SYN005	2026-07	OK	OK	95.8%	OK
SYN005	2026-08	OK	Suppressed above results	120.0%	Review
SYN005	2026-09	OK	Suppressed above results	117.6%	Review
SYN006	2026-07	OK	OK	93.6%	OK

Figure 6. The Consistency_Checks worksheet, showing the initial rows and review flags for SYN005 in August and September. All data are synthetic.

7. Clinic Visits and Clients Who May Look Lost to Follow-Up

Clinic visit reporting is important for monitoring treatment engagement, supporting medicine forecasting and identifying clients who may have interrupted treatment (IIT). Inaccurate visit counts can affect programme decisions and the follow-up of clients who may need support. A simple example:

Clients who came for a refill	In the register	In the report	Gap
One facility, one month	20	15	5 clients missing

- **When visits are under-reported,** the missing clients look as if they did not come. Programmes usually treat a client as interrupted in treatment when a month or more passes

without a visit or refill (the exact rule depends on the programme). So clients who are taking their drugs can be wrongly counted as lost, and staff may waste time tracing them.

- **Drug planning suffers.** Drug orders are based on reported visits. If visits are too low, too few drugs may be planned and a clinic can run short.
- **When visits are over-reported,** too many drugs are planned, and clients who really stopped coming can be hidden from the IIT list, so nobody follows them up.

In the workbook, 4 clinic visit records were under-reported and 1 was over-reported. Altogether 505 refill visits were left out of reports (SYN004 about 327 over three months and SYN008 about 178 in September) and 70 visits were added by over-reporting at SYN001 in August.

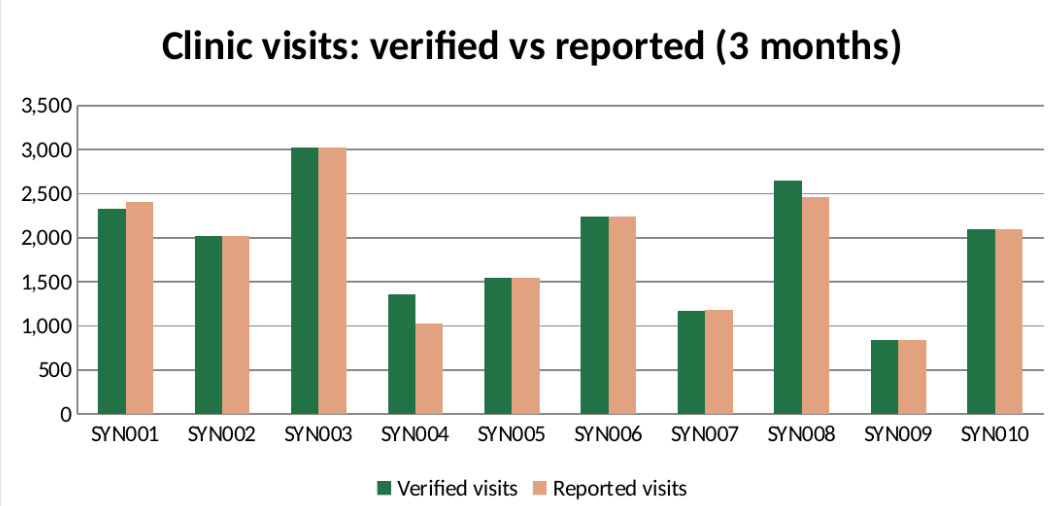


Figure 7. Clinic visits over three months: verified vs reported, by facility (synthetic data).

E. Clinic visits (drug refills) and clients who may look lost to follow-up							
Facility	Verified visits (3 months)	Reported visits (3 months)	Gap (verified minus reported)	Months under-reported	Months over-reported	Visits left out of reports	Visits added to reports
SYN001	2,335	2,405	-70	0	1	0	70
SYN002	2,024	2,024	0	0	0	0	0
SYN003	3,024	3,024	0	0	0	0	0
SYN004	1,361	1,034	327	3	0	327	0
SYN005	1,543	1,543	0	0	0	0	0
SYN006	2,240	2,240	0	0	0	0	0
SYN007	1,175	1,183	-8	0	0	0	0
SYN008	2,645	2,467	178	1	0	178	0
SYN009	846	846	0	0	0	0	0
SYN010	2,097	2,097	0	0	0	0	0
All	**19,290**	**18,863**	**427**	**4**	**1**	**505**	**70**

Figure 8. Clinic visit table from the DQA_Summary sheet (synthetic data).

8. Why Did the Problems Happen?

For each flagged record, I considered possible process-related explanations for the observed discrepancy. Five recurring causes were used to illustrate how data-quality problems can arise and what controls could help prevent them.

Cause	What it looks like in practice	Seen at
No check before reporting	Register totals are not compared with reported figures before submission.	SYN003
Registers updated late	Records are updated in batches, increasing the risk of omissions and delayed reporting.	SYN005, SYN009
Refill visits not recorded	Clients who attend only to collect medicines are omitted from the visit register.	SYN004, SYN008
Typing and adding-up mistakes	Figures are copied incorrectly, totals are calculated incorrectly or records are counted more than once.	SYN006, SYN008
Equipment problems	A faulty laptop delays data entry or database updates and may contribute to late submissions.	SYN009

These explanations are illustrative hypotheses based on the patterns in the synthetic dataset. They have not been confirmed through interviews, source-document reviews or actual facility investigations.

The examples highlight the importance of reliable reporting routines, clear recording procedures, independent checks and timely equipment support. Addressing weaknesses in processes and systems can reduce errors without placing blame on individual staff.

9. Findings and Action Log

Each finding has an ID, a likely cause, a corrective action, an owner and a status. The full log is on the DQA_Findings sheet. This turns findings into a plan that can be followed up each month.

ID	Finding	Likely cause	Action	Status
F001	SYN003: patients on treatment under-reported by about 7%, all three months	Adding-up or copying mistake	Check register against report; second-person review	In progress
F002	SYN004: clinic visits under-reported by about 24%, all three months	Refill-only visits not in the register	Match visit register to pharmacy refills; check before calling anyone IIT	Open
F003	SYN005: viral load samples and results under-reported; suppressed above results	Records not updated on time	Update through the month; weekly check with lab forms	Open
F004	SYN006: new patients over-reported by about 25% in Aug and Sep	Possible duplicate entries	Run a duplicate check before counting	Open
F005	SYN008: results above samples; clinic visits under-reported in Sep	Mistakes when adding up	Add a validation rule; second-person review	Open
F006	SYN009: values missing in Sep, under-reporting, late in Aug and Sep	Faulty laptop caused a backlog	Report faults early; IT repairs quickly; submit a corrected report	In progress
F007	SYN001: clinic visits over-reported by about 9% in Aug	Possible double counting	Count one visit per client per day	Open

ID	Finding	Likely cause	Action	Status
F008	SYN002, SYN007: small same-direction differences in all months	A steady habit or counting rule	Review trends; agree one counting rule	Open

10. Recommendations

Based on the patterns identified in the synthetic dataset, I developed the following recommendations to improve reporting accuracy, completeness, timeliness and consistency. The proposed responsibilities and timelines illustrate how findings can be translated into a practical improvement plan.

#	Recommendation	Who	When
1	Check the register against the monthly report before it is sent.	Facility data officer	Every month
2	Compare the clinic visit register with pharmacy refill records, and fix differences before reporting.	Facility data officer and pharmacist	Every month
3	Before counting any client as lost to follow-up (IIT), check that they did not collect a refill that was not reported.	Facility in-charge	Every month
4	Update registers and databases through the month, not all at month end.	Facility data officer	Every month
5	Add a second-person check before each report is submitted.	Facility in-charge	Every month
6	Agree a weekly pick-up of viral load results from the lab.	Lab focal person with clinic staff	Start next month
7	Send a reminder three days before the reporting deadline.	M&E officer	Every month
8	Add simple check formulas so totals cannot break the rules (results cannot exceed samples; suppressed cannot exceed results).	M&E officer	Within one month
9	Report laptop or software faults early, and have IT fix them quickly. Keep a backup way to report.	Facility staff and IT	Ongoing
10	Share one approved tally tool and give short refresher training based on the gaps found.	Programme manager	Within two months
11	Review trends, not only single flags, so steady same-direction differences are caught.	SI team	Every month
12	Keep an action log and review open actions until they are closed.	M&E officer	Every month
13	Give extra mentoring to facilities that have the same problem again and again.	Programme team	Every quarter
14	Repeat the DQA each quarter and track whether flags go down.	Programme team	Every quarter

Pre-submission data-quality checklist

Before submitting a monthly report, the responsible team should confirm that:

- Every facility and indicator is filled in (nothing left blank by mistake).
- Numbers match the register.
- Clinic visits match the pharmacy refill records.
- Totals make sense (for example, suppressed is not more than results received).
- A second person has looked at it.
- It is sent before the deadline.

11. Sharing the Findings

Busy programme teams rarely read long reports. So I put the main points on a short PowerPoint deck: what we found, why it matters for clinic visits and IIT, and what we recommend.

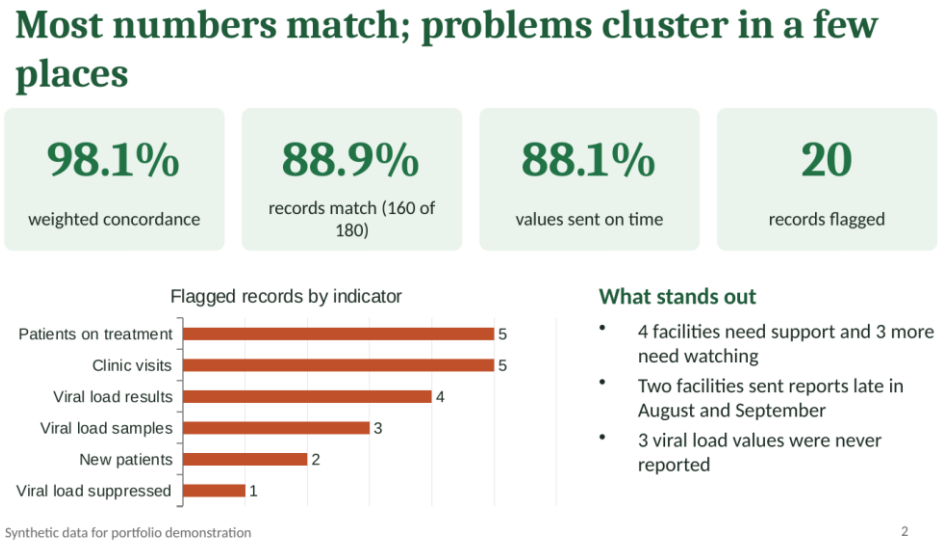


Figure 9. PowerPoint slide: headline findings (synthetic data).

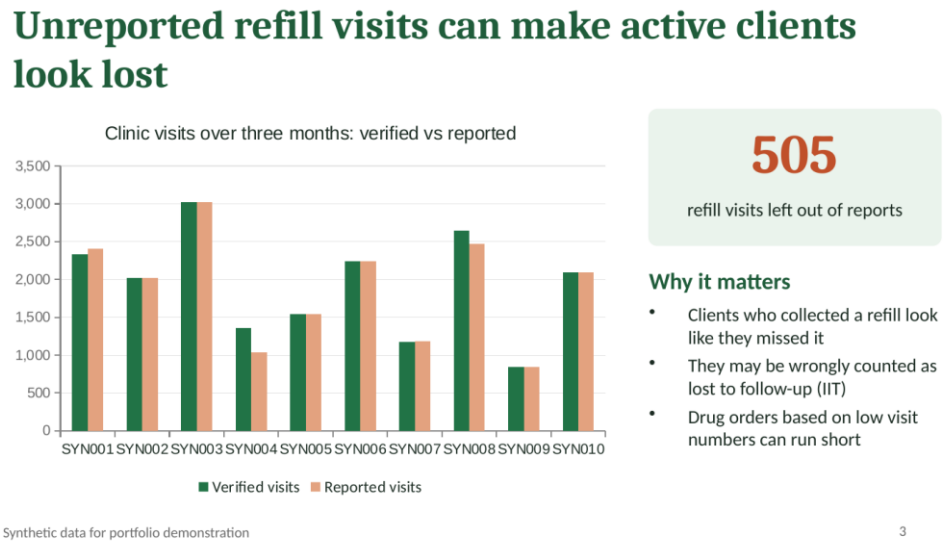


Figure 10. PowerPoint slide: unreported refill visits and clients who may look lost to follow-up (synthetic data).



Figure 11. PowerPoint slide: causes and recommended actions.

12. Limitations

- The dataset, facility codes and findings are synthetic and were built to show my method. The problems were put in deliberately, so the rates are not programme results.
- No patient-level records or real facility names were used.
- The 5% tolerance and the deadline of the 15th of the next month are assumptions chosen for this exercise.
- The IIT explanation is simplified. Real programmes use their own definitions.
- The findings are examples of how a DQA works and must not be read as results from a real assessment.

Skills Demonstrated

- Checking data quality and finding differences between registers and reports
- Looking at completeness, timeliness, accuracy and consistency as separate measures
- Linking data quality to drug planning and to clients who may look lost to follow-up (IIT)
- Finding the causes of problems and planning corrective actions with owners and dates
- Excel: formulas (SUMIFS, COUNTIFS), data validation, conditional formatting, summaries and charts
- PowerPoint: presenting findings simply for programme teams
- Explaining technical work in plain language

Files in This Portfolio

File	Contents
HIV_DQA_Workbook.xlsx	Dataset, data dictionary, consistency checks, summary, findings and dashboard
HIV_DQA_Findings_Deck.pptx	Short PowerPoint deck of findings and recommendations