

Data Quality Assessment Report

Organization: MedTrack Ghana

System: Patient Appointment & Billing Database

Role: Junior Developer

Assessment Date: November 2025

Identified Data Quality Issues

Below are **specific examples** from the dataset, mapped to **each data quality dimension**.

1. Accuracy (Factually Incorrect Data)

Example

- PhoneNumber = 244789012 (missing leading 0)
- Ghanaian mobile numbers should be **10 digits** and typically start with 0.

Why this is inaccurate

- The number does not represent a reachable Ghanaian phone number.

2. Completeness (Missing Required Information)

Example

- P004,, 0555234567, 2025-10-17, Dr. Mensah, Paid
- PatientName is missing.

Why this is incomplete

- Patient name is required for identification, reporting, and clinical context.

3. Consistency (Same Data Represented Differently)

Examples

- Dr. Osei vs dr. osei
- Paid vs paid
- Date formats:
 - 2025-10-15
 - 15/10/2025
 - 10/16/2025

Why this is inconsistent

- Same logical values are stored in different formats and casing.
-

4. Timeliness (Wrongly Dated Information)

Example

- Mixed date formats cause ambiguity:
 - 10/16/2025 → Is this **October 16** or **16th October**?

Why this is a timeliness issue

- Appointment reminders and billing depend on correct interpretation of dates.
 - A wrongly parsed date can cause reminders to be sent late or not at all.
-

5. Validity (Doesn't Follow Format Rules)

Examples

- Phone numbers:
 - 244789012 (invalid format)
- Dates:
 - 15/10/2025 (not ISO format)

Why this is invalid

- System expects predefined formats (e.g., ISO YYYY-MM-DD).
 - Invalid data may fail validation checks or processing scripts.
-

6. Uniqueness (Duplicate Records)

Example

- P001, Kwame Mensah, 0244123456 appears **twice** with different appointment dates.
- P002, Ama Serwa appears twice for the same appointment date.

Why this violates uniqueness

- Patient IDs should uniquely identify patients.
- Duplicate records inflate reports and confuse billing.

Business Impact Assessment

Data Issue	Operational Problem	Business Function Affected
Inaccurate phone numbers	SMS reminders fail	Operations
Missing patient names	Incomplete clinical records	Clinical
Inconsistent doctor names	Reports show wrong doctor workload	Operations
Ambiguous dates	Reminders sent at wrong time	Operations
Invalid formats	System rejects records	IT / Operations
Duplicate patients	Double billing, inflated reports	Finance

Recommended Solutions (Top 3 Critical Issues)

Issue 1: Invalid & Inaccurate Phone Numbers (HIGH IMPACT)

Technical Solution

- Add validation rules:
 - Must be exactly 10 digits
 - Must start with 0
- Implement a preprocessing script to normalize numbers.

Responsible Role

- Backend Developer / Data Engineer

Verification

- Run a query to confirm:
 - $\text{LENGTH(phone_number)} = 10$
 - SMS delivery success rate improves

Issue 2: Inconsistent Date Formats (HIGH IMPACT)

Technical Solution

- Enforce ISO date format (YYYY-MM-DD) at data entry.
- Convert legacy dates using a standardization script.

Responsible Role

- Backend Developer
- QA Engineer for validation testing

Verification

- All appointment dates parse correctly
- Reminder jobs trigger exactly 24 hours before appointments

Issue 3: Duplicate Patient Records (HIGH IMPACT)

Technical Solution

- Enforce PatientID as a **primary key**
- Run deduplication using:
 - PatientID + PhoneNumber matching

Responsible Role

- Database Administrator (DBA)

Verification

- Patient count matches unique PatientIDs
- Billing records no longer duplicate charges

4. Biggest Risk of Poor Data Consistency (From a Developer / Systems Perspective)

The **biggest risk** is **automation failure**.

When data is inconsistent:

- Validation rules break
- Reports become unreliable
- Automated workflows (SMS reminders, billing, analytics) fail silently

In health tech, this directly affects **patient trust, clinical efficiency, and revenue integrity**.

5. Conclusion

This assessment identified violations across **all six data quality dimensions**, each directly linked to MedTrack Ghana's operational issues. Addressing validation, standardization, and deduplication will significantly improve system reliability, patient communication, and financial accuracy.

