## The Facility Data Quality Assessment 2024

**Purpose of the Query**

The [facility scoring query](https://adb-2733269684173368.8.azuredatabricks.net/editor/notebooks/762023208272404?o=2733269684173368#command/8042436903824214) is designed to evaluate the data completeness and timeliness of facilities based on six critical metrics. These metrics focus on both the quality and accuracy of the data submitted by facilities, encompassing areas such as timeliness of data entry, completeness of demographic information, and accuracy of key operational fields like funding source and lot numbers. By assessing performance against these six metrics, the query assigns each facility a score ranging from 0 to 12. Points are awarded based on defined thresholds that categorize performance as excellent, average, or needing improvement.

The primary goal of this scoring system is to identify facilities that excel in maintaining high-quality, complete, and timely data while also highlighting those that require targeted support to address data gaps. By standardizing the evaluation process, the scoring query provides an objective framework to measure performance and facilitates consistent comparisons across facilities. Additionally, this system empowers stakeholders to make data-driven decisions by identifying specific areas where resources and interventions can be most effectively applied. Ultimately, this contributes to the improvement of data quality at the facility level, ensuring the reliability of health records and supporting broader public health objectives.

**Metrics Evaluated**

The scoring system evaluates facilities across the following six metrics:

1. **Timeliness**:

Measures the percentage of records entered more than one day (before or) after the vaccination date.

1. **Missing VFC Eligibility**:

Calculates the percentage of records missing the VFC eligibility field.

1. **Missing Funding Source**:

Evaluates the percentage of records lacking funding source information.

1. **Missing Lot Number**:

Identifies the percentage of records with missing or invalid lot numbers.

1. **Missing Race**:

Assesses the percentage of records with no valid race information recorded.

1. **Missing Ethnicity**:

Evaluates the percentage of records missing valid ethnicity data.

Each metric has clearly defined thresholds that determine the points awarded (0, 1, or 2) based on facility performance. These thresholds are designed to differentiate facilities that excel in data quality from those that show average performance or need significant improvement. Lower percentages within each metric represent the ideal range of performance, reflecting adherence to high standards of data timeliness and completeness. Mid-range percentages are aligned with typical averages observed across facilities, providing a baseline for acceptable performance. Higher percentages, on the other hand, indicate substantial gaps in data quality, leading to lower scores and signaling areas that require immediate attention. By establishing these thresholds, the scoring system ensures an objective and consistent evaluation process, allowing stakeholders to easily identify both exemplary facilities and those in need of targeted support.

**Query Breakdown**

1. **Filtered Data**:

The query begins by filtering vaccination records from 2024.This ensures only relevant data is evaluated.

1. **Subqueries**:
   * **Timeliness**:
     + The subquery calculates timeliness using vaccination\_mater
     + It counts all records (COUNT(\*) AS Total\_Records) and calculates the percentage of records entered more than one day after the vaccination date (ROUND(100.0 \* COUNT(CASE WHEN DATEDIFF(DAY, VM.VACC\_DATE, VM.INSERT\_STAMP) > 1 THEN 1 ELSE NULL END) / COUNT(\*), 2) AS Beyond\_One\_Day\_Percent).
     + This subquery is joined to facility\_master (aliased as FM) using FM.ASIIS\_FAC\_ID = T.ASIIS\_FAC\_ID.
     + **Thresholds**:
       - ≤ 5%: 2 points
       - > 5% and ≤ 10.17%: 1 point
       - > 10.17%: 0 points
   * **Missing VFC Eligibility**:
     + The percentage of records with missing VFC Eligibility is calculated using vaccination\_mater table
     + The numerator counts records where VFC\_ELIGIBLE is NULL, while the denominator is the total record count (ROUND(100.0 \* COUNT(CASE WHEN VM.VFC\_ELIGIBLE IS NULL THEN 1 ELSE NULL END) / COUNT(\*), 2) AS Missing\_VFC\_Eligible\_Percent).
     + Joined to facility\_master using FM.ASIIS\_FAC\_ID = V.ASIIS\_FAC\_ID.
     + **Thresholds**:
       - ≤ 1%: 2 points
       - > 1% and ≤ 2.87%: 1 point
       - > 2.87%: 0 points
   * **Missing Funding Source**:
     + This subquery uses VACCINATION\_MASTER (VM) data to calculate the percentage of missing funding source data.
     + The numerator counts records missing a FUNDING\_SOURCE\_ID, while the denominator is the total record count (ROUND(100.0 \* COUNT(CASE WHEN VM.FUNDING\_SOURCE\_ID IS NULL THEN 1 ELSE NULL END) / COUNT(\*), 2) AS Missing\_Funding\_Source\_Percent).
     + Joined to facility\_master using FM.ASIIS\_FAC\_ID = F.ASIIS\_FAC\_ID.
     + **Thresholds**:
       - ≤ 1%: 2 points
       - > 1% and ≤ 8.91%: 1 point
       - > 8.91%: 0 points
   * **Missing Lot Number**:
     + This metric uses VACCINATION\_MASTER (VM) and checks against lot\_number (LN).
     + The numerator counts records with missing or invalid lot numbers (ROUND(100.0 \* COUNT(CASE WHEN VM.LOT\_NUM IS NULL OR NOT EXISTS (SELECT 1 FROM lot\_number AS LN WHERE VM.LOT\_NUM = LN.LOT\_NUMBER) THEN 1 ELSE NULL END) / NULLIF(COUNT(\*), 0), 2) AS Missing\_Lot\_Number\_Percent).
     + Joined to facility\_master using FM.ASIIS\_FAC\_ID = L.ASIIS\_FAC\_ID.
     + **Thresholds**:
       - ≤ 1%: 2 points
       - > 1% and ≤ 18.74%: 1 point
       - > 18.74%: 0 points
   * **Missing Race**:
     + This metric uses VACCINATION\_MASTER (VM) joined with PATIENT\_MASTER (PM) and PATIENT\_RACE\_RESERVE (PRR).
     + The numerator counts records where no valid race exists in the reserve table, and the denominator is the total count of vaccination records (ROUND(100.0 \* SUM(CASE WHEN NOT EXISTS (SELECT 1 FROM PATIENT\_RACE\_RESERVE PRR WHERE PRR.ASIIS\_PAT\_ID\_PTR = PM.ASIIS\_PAT\_ID AND PRR.PAT\_RACE <> '9') THEN 1 ELSE 0 END) / COUNT(\*), 2) AS Missing\_Race\_Percent).
     + Joined to facility\_master using FM.ASIIS\_FAC\_ID = R.ASIIS\_FAC\_ID.
     + **Thresholds**:
       - ≤ 1%: 2 points
       - > 1% and ≤ 3.68%: 1 point
       - > 3.68%: 0 points
   * **Missing Ethnicity**:
     + This subquery uses VACCINATION\_MASTER (VM) joined with PATIENT\_MASTER (PM) and PATIENT\_RESERVE (PR).
     + The numerator counts records without valid ethnicity codes (ROUND(100.0 \* SUM(CASE WHEN NOT EXISTS (SELECT 1 FROM PATIENT\_RESERVE PR WHERE PR.ASIIS\_PAT\_ID\_PTR = PM.ASIIS\_PAT\_ID AND PR.PAT\_ETHNICITY\_CODE IN (1, 2)) THEN 1 ELSE 0 END) / COUNT(\*), 2) AS Missing\_Ethnicity\_Percent).
     + **Joins**: Joined to facility\_master using FM.ASIIS\_FAC\_ID = E.ASIIS\_FAC\_ID.
     + **Thresholds**:
       - ≤ 1%: 2 points
       - > 1% and ≤ 6.14%: 1 point
       - > 6.14%: 0 points

**Score Calculation**

Points are assigned using conditional statements ("CASE WHEN") based on predefined thresholds that categorize performance into three tiers: excellent, average, or needing improvement. These thresholds are carefully determined to align with realistic performance goals and average benchmarks across facilities.

The total score for each facility is calculated by summing the points earned across all six metrics. This cumulative score, ranging from 0 to 12, provides an overall indicator of the facility’s data quality. Facilities scoring closer to 12 demonstrate strong performance across all evaluated areas, while lower scores highlight specific metrics where improvements are needed.

**Result and Summery**

The facility scoring system evaluates data completeness and timeliness using six critical metrics, assigning scores from 0 to 12. These metrics cover key areas such as timeliness of data entry, demographic information completeness, and accuracy of operational fields like funding source and lot numbers. Performance is categorized into four tiers—excellent, good, fair, and needing improvement—based on predefined thresholds.

* **High-Scoring Facilities (Scores: 10–12)**:
  + **434 facilities** fall into this category, showcasing excellent data quality, timeliness, and completeness.
  + These facilities likely have strong processes for maintaining accurate demographic and operational data.
  + They serve as models of best practices, providing a benchmark for others.
* **Good-Scoring Facilities (Scores: 8–9)**:
  + **799 facilities** fall into this range, demonstrating solid performance with some minor areas for improvement.
  + These facilities generally maintain good practices in data completeness and accuracy.
* **Fair-Scoring Facilities (Scores: 5–7)**:
  + **1,045 facilities** are in this range, representing average performance with moderate gaps in certain areas.
  + These facilities require improvements in timeliness, demographic completeness, or operational fields.
* **Low-Scoring Facilities (Scores: 0–4)**:
  + **693 facilities** fall into this category, highlighting significant challenges in data quality, completeness, and timeliness.
  + They are often plagued by substantial gaps in operational accuracy, such as funding sources or lot numbers, and missing demographic information.
  + These facilities indicate a need for focused support and resources to address systemic issues and improve performance.

**Recommendations**

1. **Recognize High Performers**:
   * Share successful practices from the 434 high-scoring facilities to encourage peer learning.
   * Consider engaging these facilities in mentoring or knowledge-sharing roles.
2. **Monitor Good and** **Fair-Scoring Facilities**:
   * Focus on the 1,844 mid-scoring facilities by identifying specific data gaps and providing customized support.
   * Offer training on data entry processes and operational accuracy improvements.
3. **Support Low-Scoring Facilities**:
   * Prioritize interventions for the 693 low-scoring facilities.
   * They are likely in need for targeted support to bridge performance gaps.

**Targeted Support**

The **Low-Scoring Facilitie**s include facilities scoring between 0 and 4, indicating significant challenges in data quality, completeness and timeliness. Below is an analysis based on facility names, total records, and key performance metrics:

**Key Observations**

1. **Performance Challenges**:
   * **Timeliness**: High percentages of records exceed acceptable entry timelines, with delays such as over 30% in some facilities.
   * **Missing Operational Data**:
     + **Missing Funding Source**: Many facilities exhibit significant gaps.
     + **Missing Lot Numbers**: Over 35% of data is incomplete in some cases.
   * **Demographics**: Missing race and ethnicity data often exceed 10–20% in these facilities.
2. **Facility Size**:
   * Facilities processing fewer records tend to show more challenges. For instance, facilities with fewer than 5000 records show higher percentages of missing data.
3. **Facility Types**:
   * Facilities with names referencing specific towns or communities (e.g., "XXXXXX COMMUNITY HEALTH CENTER") may indicate rural or resource-limited locations.
   * Specialized or temporary facilities (e.g., "Pediatrics Unit" or "Mobile Health Clinic") often prioritize care delivery, which can result in administrative gaps.

**Next Steps**

** Contact Low-Scoring Facilities Directly:**

* Establish a direct communication channel with facilities scoring between 0 and 4.
* Schedule meetings to understand specific barriers, such as staffing shortages, technical issues, or workflow inefficiencies.
* Provide tailored support plans for each facility, focusing on their unique challenges.
* Offer follow-up sessions to monitor progress and provide additional resources if needed.

** Focus on Rural and Smaller Clinics:**

* Provide targeted training on data quality and reporting processes.
* Introduce simplified workflows or automation tools for facilities with limited staff.

** Improve Timeliness:**

* Implement alerts or reminders for timely data entry.
* Support facilities with streamlined systems to reduce delays.

** Enhance Demographic Data Collection:**

* Provide resources to standardize intake processes.
* Introduce tools to reduce manual errors and ensure demographic completeness.

** Engage High-Need Facilities:**

* Prioritize regular follow-ups with low-performing facilities to ensure sustained improvement.
* Offer technical assistance programs that include on-site or virtual training, troubleshooting, and data quality audits.

** Monitor and Measure Progress:**

* Run this analysis regularly and establish performance benchmarks for low-scoring facilities and track their progress.
* Use other data analytics to identify trends and measure the effectiveness of interventions.

This analysis underscores the critical need for targeted support to bridge performance gaps within these facilities. Facilities with poor scores often face systemic challenges, including limited staffing, resource constraints, and outdated processes, all of which contribute to delays and inaccuracies in data reporting. By addressing these issues, stakeholders can strengthen the overall quality of health data, which is essential for effective decision-making and resource allocation. By closing data gaps in timeliness, demographic completeness, and operational accuracy, these facilities can better align with broader public health objectives, such as improving disease surveillance, promoting health equity, and enhancing the overall quality of healthcare data.

Ultimately, addressing these challenges will have a cascading effect: empowering facilities to operate more effectively, improving patient care outcomes, and ensuring the integrity of health systems at all levels. Targeted interventions, coupled with ongoing support and monitoring, will help create a sustainable framework for continuous data quality improvement, ensuring that all facilities contribute meaningfully to the public health ecosystem.