OSIIS Data Quality

Martin Lansdale, OSIIS Data Quality Manager
martinl@health.ok.gov
Today we will discuss:

- What is Data Quality
- IIS Data Quality Blue Print
- Why Does Data Quality Matter
- OSIIS Data Quality Plan
- OSIIS Data Quality Review
- OSIIS Data Quality Improvement
Data Quality: What is it?

• Data Quality is an integral part of data governance that ensures your organizations data (i.e. IIS registry) is fit for its purpose.

• Generally refers to the overall utility of a dataset and includes 6 core concepts: Accuracy, Completeness, Consistency, Timeliness, Validity, and Uniqueness.

• All core concepts are important but an organization can focus on specific ones depending on need.

• The CDC, MIROW, and AIRA have created a ‘best practices” guide that includes a IIS Data Quality Blue Print which focuses on 4 specific core concepts.
DATA QUALITY: WHAT IS IT?

DATA QUALITY CHARACTERISTICS

As mentioned, the best practice recommendations made in the guide support the Immunization Information Systems (IIS) Data Quality Blueprint.

MAIN CHARACTERISTICS

- **Availability:** IIS data can be readily used by authorized users to inform IIS and immunization program activities.

- **Completeness:** The IIS contains all (historic and new) vaccination and demographic records for persons residing in the jurisdiction, and all vaccination and demographic records contain complete data fields.

- **Timeliness:** Patient records are established, and vaccination events are recorded in the IIS within specified time frames.

- **Validity:** Vaccination and demographic records in the IIS conform with generally accepted standards (e.g., Advisory Committee on Immunization Practices recommendations, United States Postal Service standards).
The IIS Data Quality Blueprint is a guide to help immunization programs address and advance data quality within IISs. Its measures focus on improving IIS data throughout the lifespan.

The blueprint focuses on driving IIS data that are:
- AVAILABLE
- COMPLETE
- TIMELY
- VALID

Other enabling characteristics (i.e., ACCURATE, CONSISTENT, and UNIQUE) and functional processes (e.g., interoperability) are inherent to high quality IIS data and complement the blueprint data quality characteristics.

Sample activities for achieving high quality data and the associated recommended measures are found in the IPOM Sample Activities 2023 document.
“You can have all of the fancy tools, but if [your] data quality is not good, you're nowhere.” — Veda Bawo, director of data governance, Raymond James

“Without clean data, or clean enough data, your data science is worthless.” — Michael Stonebraker, adjunct professor, MIT
• Data Quality ensures that data is fit for its purpose and can accurately and reliably be used for analysis.
• Poor data quality can lead to erroneous associations causing bad policy/decision making.
• Measuring data quality can help organizations identify data gaps/errors that need to be resolved.
• Mantra: Crap in, Crap out. If your data is unreliable then your analysis is suspect or at least reduced in scope.
OSIIS Data Quality Plan
OSIIS Data Quality Plan

- Completed in June 2022 in conjunction with OSDH and AIRA staff.
- Based on IIS Data Quality Blueprint and MIROW guidelines.
- SCOPE: Improving the quality of data at rest in OSIIS.
- The scope includes assessment of patient demographic and vaccination information but not a review of incoming data.
- Plan includes 8 Objectives and strategies to improve data quality.
- Objectives include 6 high priority issues.
• Lack of data exchange with Vital Records: Need Vital Records feed for birth and death records.

• Missing jurisdiction-level patient status: Jurisdiction-level patient status is important in order to appropriately include or exclude individuals from vaccination coverage assessments.

• Cost of maintaining storage space for legacy OSIIS data.

• Incomplete phone number/email data in patient records.

• Incomplete race and ethnicity data in patient records.

• Errors in vaccine funding source in vaccination records.
OSIIS Data Quality Review
• Assessed 2021 and 2022 data for all data elements in the Oklahoma Data Quality Plan.

• A total of 9,610,838 records reviewed, core concept: Completeness.

• Current administrations and history shots separated.

• Manual Entry and HL7 submission compared along with patient age.

• Veterans Health Administration (VHA) and IZ Gateway data compared.

• Analysis was completed with SAS v9.4 and Excel.
OSII S Data Quality Review: History vs. Current

OSII S Data Quality: History vs Current Administration

- **2021 Year**
  - Manual Submission: 93.8% Current, 6.2% History
  - HL7 Submission: 94.1% Current, 5.9% History
  - OOS/VHA HL7 Submission: 91.7% Current, 8.3% History

- **2022 Year**
  - Manual Submission: 89.8% Current, 10.3% History
  - HL7 Submission: 91.7% Current, 8.3% History
  - OOS/VHA HL7 Submission: 89.7% Current, 10.3% History

- **Overall**
  - Manual Submission: 91.9% Current, 8.1% History
  - HL7 Submission: 93.1% Current, 6.9% History
  - OOS/VHA HL7 Submission: 91.9% Current, 8.1% History
OSIIS Data Quality Review: Submission Status

OSIIS Data Quality by Submission Type, 2021-2022

- Manual Submission: 22.77% (2021) vs 28.26% (2022)
- HL7 Submission: 77.21% (2021) vs 71.51% (2022)
- OOS/VHA HL7 Submission: 0.02% (2021) vs 0.23% (2022)
- Total Shot Percentage: 66.42% (2021) vs 33.58% (2022)

*28.7% (n=1,130) of providers in OSIIS submitted through HL7
OSIIS DATA QUALITY REVIEW: FUNDING SOURCE

OSIIS Shot Administrations by Funding Source, 2019-2022

* Covid Vaccine Roll-Out. PAN funding source used for covid shots only.

<table>
<thead>
<tr>
<th>Year</th>
<th>317</th>
<th>PAN</th>
<th>PRIVATE</th>
<th>STATE</th>
<th>UNKNOWN</th>
<th>VFC</th>
</tr>
</thead>
<tbody>
<tr>
<td>2019</td>
<td>1.3%</td>
<td>0.0%</td>
<td>20.9%</td>
<td>3.1%</td>
<td>0.2%</td>
<td>74.5%</td>
</tr>
<tr>
<td>2020</td>
<td>0.6%</td>
<td>6.0%</td>
<td>20.7%</td>
<td>1.8%</td>
<td>0.2%</td>
<td>70.5%</td>
</tr>
<tr>
<td>2021*</td>
<td>0.2%</td>
<td>70.7%</td>
<td>14.0%</td>
<td>0.4%</td>
<td>0.8%</td>
<td>14.0%</td>
</tr>
<tr>
<td>2022</td>
<td>0.5%</td>
<td>27.8%</td>
<td>41.3%</td>
<td>0.8%</td>
<td>1.5%</td>
<td>28.1%</td>
</tr>
<tr>
<td>Total (4 Year)</td>
<td>0.4%</td>
<td>44.8%</td>
<td>22.6%</td>
<td>0.9%</td>
<td>0.9%</td>
<td>30.4%</td>
</tr>
</tbody>
</table>
### OSIIS Data Quality Review: Lowest Missing Percentage

<table>
<thead>
<tr>
<th>Data Element</th>
<th>OSIIS Overall % missing</th>
<th>MIROW Guideline</th>
<th>Meets?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient First Name</td>
<td>0%</td>
<td>0%</td>
<td>Yes</td>
</tr>
<tr>
<td>Patient Last Name</td>
<td>0%</td>
<td>0%</td>
<td>Yes</td>
</tr>
<tr>
<td>Patient Date of Birth</td>
<td>0%</td>
<td>0%</td>
<td>Yes</td>
</tr>
<tr>
<td>Patient Gender</td>
<td>0%</td>
<td>5%</td>
<td>Yes</td>
</tr>
<tr>
<td>Patient Address</td>
<td>1.08%</td>
<td>15%</td>
<td>Yes</td>
</tr>
<tr>
<td>Patient City</td>
<td>1.10%</td>
<td>15%</td>
<td>Yes</td>
</tr>
<tr>
<td>Patient County</td>
<td>2.02%</td>
<td>15%</td>
<td>Yes</td>
</tr>
<tr>
<td>Patient State</td>
<td>1.08%</td>
<td>15%</td>
<td>Yes</td>
</tr>
<tr>
<td>Patient Zip Code</td>
<td>1.10%</td>
<td>15%</td>
<td>Yes</td>
</tr>
<tr>
<td>Shot Manufacturer</td>
<td>5.78%</td>
<td>1%</td>
<td>No</td>
</tr>
<tr>
<td>Lot Number</td>
<td>1.15%</td>
<td>1%</td>
<td>Yes</td>
</tr>
<tr>
<td>Site</td>
<td>6.69%</td>
<td>10%</td>
<td>Yes</td>
</tr>
<tr>
<td>Route</td>
<td>5.71%</td>
<td>10%</td>
<td>Yes</td>
</tr>
<tr>
<td>Expiration date</td>
<td>5.72%</td>
<td>10%</td>
<td>Yes</td>
</tr>
<tr>
<td>Eligibility</td>
<td>5.69%</td>
<td>1%</td>
<td>No</td>
</tr>
<tr>
<td>Funding Source</td>
<td>6.75%</td>
<td>0%</td>
<td>No</td>
</tr>
</tbody>
</table>
OSIIS DATA QUALITY REVIEW: RACE

Data Quality: Racial Category by Age and Year

*Valid Race means an actual racial category was picked (not missing, blank or Other).

Data Quality: Racial Category by Submission Status and Year

*Valid Race means an actual racial category was picked (not missing, blank or Other).
OSIIS DATA QUALITY REVIEW: ETHNICITY

Data Quality: Ethnicity by Age and Year

*Valid Ethnicity means an actual ethnicity was picked (not missing/blank or unknown).

Data Quality: Ethnicity by Submission Status and Year

*Valid Ethnicity means an actual ethnicity was picked (not missing/blank or unknown).
OSIIS DATA QUALITY REVIEW: EMAIL

Data Quality: Email by Age and Year

OSIIS Public Portal
OSIIS DATA QUALITY REVIEW: PHONE NUMBER

Data Quality: Phone # by Age and Year

<table>
<thead>
<tr>
<th>Year</th>
<th>0 to 5</th>
<th>6 to 12</th>
<th>13 to 18</th>
<th>&gt; 18</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>2021</td>
<td>62.8%</td>
<td>50.7%</td>
<td>49.7%</td>
<td>88.0%</td>
<td>91.2%</td>
</tr>
<tr>
<td>2022</td>
<td>37.2%</td>
<td>24.0%</td>
<td>12.6%</td>
<td>13.5%</td>
<td>20.4%</td>
</tr>
</tbody>
</table>

Data Quality: Phone # by Submission Status and Year

<table>
<thead>
<tr>
<th>Year</th>
<th>Manual Submission</th>
<th>HLT Submission</th>
<th>OOS/VHA HLT submission</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>2021</td>
<td>51.1%</td>
<td>88.9%</td>
<td>90.5%</td>
<td>80.3%</td>
</tr>
<tr>
<td>2022</td>
<td>40.9%</td>
<td>11.1%</td>
<td>9.5%</td>
<td>26.4%</td>
</tr>
</tbody>
</table>
OSIIS DATA QUALITY REVIEW: ADMINISTERING PROVIDER

Data Quality: *Administering Provider by Age and Year

- 2021: 8.9% 78.1% 89.9%
- 2022: 8.1% 75.8% 89.9%
- 2021 Year 0 to 5: 91.1%
- 2022 Year 0 to 5: 88.9%
- 2021 Year 6 to 12: 87.7%
- 2022 Year 6 to 12: 85.8%
- 2021 Year 13 to 18: 88.9%
- 2022 Year 13 to 18: 89.9%
- 2021 Year > 18: 41.1%
- 2022 Year > 18: 21.9%

Data Quality: *Administering Provider by Submission Status

- 2021: 98.4%
- 2022: 98.6%
- Manual Submission 2021: 96.7%
- Manual Submission 2022: 98.6%
- HL7 Submission 2021: 98.4%
- HL7 Submission 2022: 96.7%
- *OSIS/PHA HL7 submission 2021: 4.1%
- *OSIS/PHA HL7 submission 2022: 9.8%
• Data Quality for certain demographic data elements is low in the younger age groups (0 to 5).
• Data Quality for providers entering manually is lower overall compared to providers submitting through HL7.
• Missing shots (VFC vs. Privately funded shots).
• Data for the actual shots given has better overall data quality than demographic elements.
• Email submission is extremely low (affects public portal) for all age groups and submission type.
OSIIS Data Quality Improvement
OSIIS Data Quality Improvement: Current Strategy

- Ensure high quality data through HL7 submission (onboarding thresholds, monthly data quality reports).
- Completing high priority objectives and strategies in data quality plan.
- Assessing data quality at the health district level and contacting providers about areas of low data quality.
- Utilizing CDC and OSDH Immunization Service staff to work with providers to fix various data quality issues.
- Engaging state OSIIS vendor to automate data quality reports for all submission types.
To ensure high data quality, multiple approaches have to be used including programmatic and technical resources.

**Educate provider organizations about data quality (make it visible).**

Prioritize data quality for data elements that have the most impact for public health and vaccine accountability.

Assess data quality throughout the process of receiving and managing data (onboarding through evaluating data at rest) for all submission types (manual and HL7).

Adoption of uniform business rules across jurisdictions that allow for some flexibility.
Questions?

Immunization/OSIIS Information