IIS DATA QUALITY PRACTICES

TO MONITOR AND EVALUATE DATA AT REST

5.11.2018

AIRA
AMERICAN IMMUNIZATION REGISTRY ASSOCIATION

EXECUTIVE SUMMARY

Data Quality Practices to Monitor and Evaluate Data at Rest is the third in a series of guides that provide practical guidance to immunization information systems (IIS) staff for assessing and improving the quality of their data. The prior two guides focused respectively on data validation during the onboarding process and ongoing data quality assessment for incoming data.

The focus for this guide is data at rest. The purpose of the guide is to provide practical guidance on techniques, methodologies, and processes for IIS to use in assessing the quality of data at rest. We define "data at rest" as data that is currently in the live, production IIS environment. The target users for this guide are IIS and immunization program staff at the state or jurisdiction level.

As in the two previous guides, this guide focuses on the data quality dimensions of completeness, accuracy, and timeliness, with the addition of validity, consistency, and uniqueness. Examining data at rest provides unique opportunities to cut and dice data across providers and to look for patterns of issues not otherwise apparent. For example, aggregated data can reveal problems with data consistency, such as unexplained changes in the volume of immunizations administered over a given period or changes in the proportions of administered vaccinations by age group. In addition, problems with vaccine coding can occur when a new vaccine is introduced and may be easier to spot when looking at aggregate

The purpose of the guide is to provide practical guidance on techniques, methodologies, and processes for IIS to use in assessing the quality of data at rest.

data. Another example, briefly addressed in the guide, concerns duplicate records and the incorrect merging of records. These may be discovered only through evaluation of data at rest from different provider sources.

To develop indicators for data quality, we used the two previous data quality guides, the *AIRA Data Quality Assurance in Immunization Information Systems: Incoming Data*, the Centers for Disease Control and Prevention (CDC) Health Level 7 (HL7) Implementation Guides (HL7 version 2.5.1, release 1.5 plus addendum), the CDC IIS Functional Standards, and documents gathered directly from IIS programs. We did a literature review that included documents from the data management community as well as reports and presentations from public health and IIS professionals. A workgroup of subject matter experts (SMEs) provided valuable input on their real-life experiences in improving IIS data quality.

The guide provides tables of indicators for evaluating data quality. The tables are organized by data quality dimension and provide a brief description of each measure and its significance. In addition, the workgroup assigned a recommended priority level for each indicator. It also assigned recommended target levels to the completeness measures. The guide offers information on a few select activities that can improve data quality at the system level. This includes strategies for cleansing and correcting addresses and for preventing duplicates and bad merges related to birth data. Finally, a section on implementation considerations gives general recommendations and a step-by-step template for building a data at rest quality analysis plan.

The usefulness of an IIS is dependent on the quality of its data. Trustworthy data is needed for clinical decision making, vaccine tracking and accountability, vaccination coverage assessments, and public health research. Recommendations made in the guide are independent of particular IIS implementations and technology solutions, and each IIS should adapt them to their own specific needs.

TABLE OF CONTENTS

TABLE OF CONTENTS, CONT.

TABLES AND FIGURES

SECTION I. INTRODUCTION

Background

> Immunization information systems (IIS) are electronic population-based health information systems that are intended to record all immunizations for patients in each geopolitical area. For the past 20 years, IIS have consolidated patient and immunization records from multiple data sources (e.g., immunization providers and Vital Statistics).

IIS also provide a diverse array of functions, such as clinical decision support (CDS) for immunizations, vaccine inventory management and accountability, reminder/recall tools, and coverage assessment reports for providers and public health agencies. Historically, immunization data have been entered manually into the IIS by clinicians, but more recently, IIS have adopted real-time electronic data exchange (EDE) between IIS and electronic health record (EHR) systems, which allows for a more automated and streamlined process.

Given the increased prominence and multiple roles of IIS in both public health and private health care systems, data quality is more important than ever. Not surprisingly, data quality continues to be named a top priority and challenge by IIS managers and staff, as described in the 2017 American Immunization Registry Association (AIRA) Education Survey.[1] Respondents chose data quality as the number one need over eight other high-priority IIS topics. Specific subtopics within this category included tools and strategies for onboarding data validation, ongoing validation of incoming data, patient and vaccine deduplication, and monitoring of data at rest. Two previously published AIRA guides focused respectively on data validation during the onboarding process and during the ongoing data submission process, hereinafter referred to as the Phase One[2] and Phase Two guides.[3] The present guide, titled *IIS Data Quality Practices to Monitor and Evaluate Data at Rest* (hereinafter referred to as the guide), completes AIRA's three-part series on data quality.

Purpose

The purpose of the guide is to provide practical guidance on techniques, methodologies, and processes for IIS to use in assessing the quality of data at rest. **Data at rest is quite simply defined as data that has been accepted into the IIS production environment.** This data includes demographic and immunization record information that is currently in the live, production environment (e.g., database or other data store).

[1] AIRA. 2017 Education Survey Summary. July 2017.
[2] AIRA. Data Validation Guide for the Onboarding Process. Sept 2016. http://repository.immregistries.org/resource/data-validation-guide-for-the-iis-onboarding-process/.
[3] AIRA. IIS Data Quality Practices: Monitoring and Evaluating Data Submissions. Sept 2017. http://repository.immregistries.org/resource/iis-data-quality-practices-monitoring-and-evaluating-data-submissions/.

Data at rest is defined as data that has been accepted into the IIS production environment.

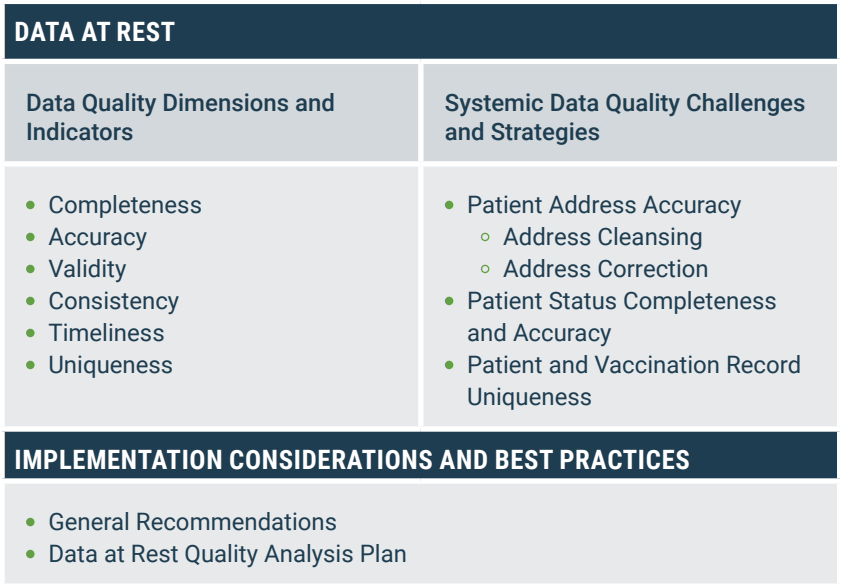
Depending on the IIS protocols in place at the time of entry, the data may have been processed with varying degrees of screening, cleansing, or consolidation. Monitoring and evaluating data at rest is also known as retrospective review.

As described in the previous two AIRA data quality guides, preventing erroneous data from entering the IIS is essential to data quality. However, despite best efforts at prevention, erroneous data can still occur within the IIS database. Bad data can result from manual data entry errors, submissions that pre-date incoming data quality rules, automated back-end processes that erroneously alter data, changes in algorithms, and coding issues. It also may result from data sources that do not undergo routine monitoring during data submission. The guide identifies and highlights proven practices used in the IIS community for data at rest quality review. The intent is to give IIS a resource to build or expand their data quality practices. It is not intended to describe a one-size-fits-all approach. Topics covered in this document include:

- A description and prioritization of data quality measures
- Systemic data issues related to address accuracy and special deduplication and record merge situations
- General implementation and process considerations
- A step-by-step guide to developing a data quality plan for data at rest
- Sample reports used by IIS programs to monitor and evaluate data at rest

A visual depicting the information contained in the guide is presented in Figure 1.

Figure 1 | *Data at Rest Diagram at a High-level View*



Scope of Guide

The guide focuses on IIS practices related to the routine monitoring and evaluation of data at rest. The three data quality dimensions of accuracy, completeness, and timeliness are addressed. The additional data quality dimensions of validity, consistency, and uniqueness are also covered. While the first two guides in this series focused on provider-specific data quality, this guide examines data that has been aggregated by other parameters. For example, back-end data investigation may look at data at a geographic level or by age group. The findings may lead to deeper analysis at the provider level, but provider-level data does not need to be the starting point.

Patient- and vaccination-level deduplication is an essential activity for every IIS, and real-time and retrospective searches for duplicate records in the IIS should be considered a universal best practice. Several resources on this topic have already been developed that define best practices, processes, and algorithms for deduplication. Although more guidance may need to be developed, the complexity of the topic and time available for the project did not allow for its inclusion in the guide. However, a brief discussion of special situations for deduplication and incorrect merges can be found in Section III. More details and resources on the topic are in Appendix C.

The data types of interest in this guide are primarily related to patient demographic and vaccination records. Also important but not covered here are data elements related to facility identification, user management, immunizing provider identification, and expired vaccines. The SME workgroup identified these as important to monitor but out of scope for this guide.

Audience

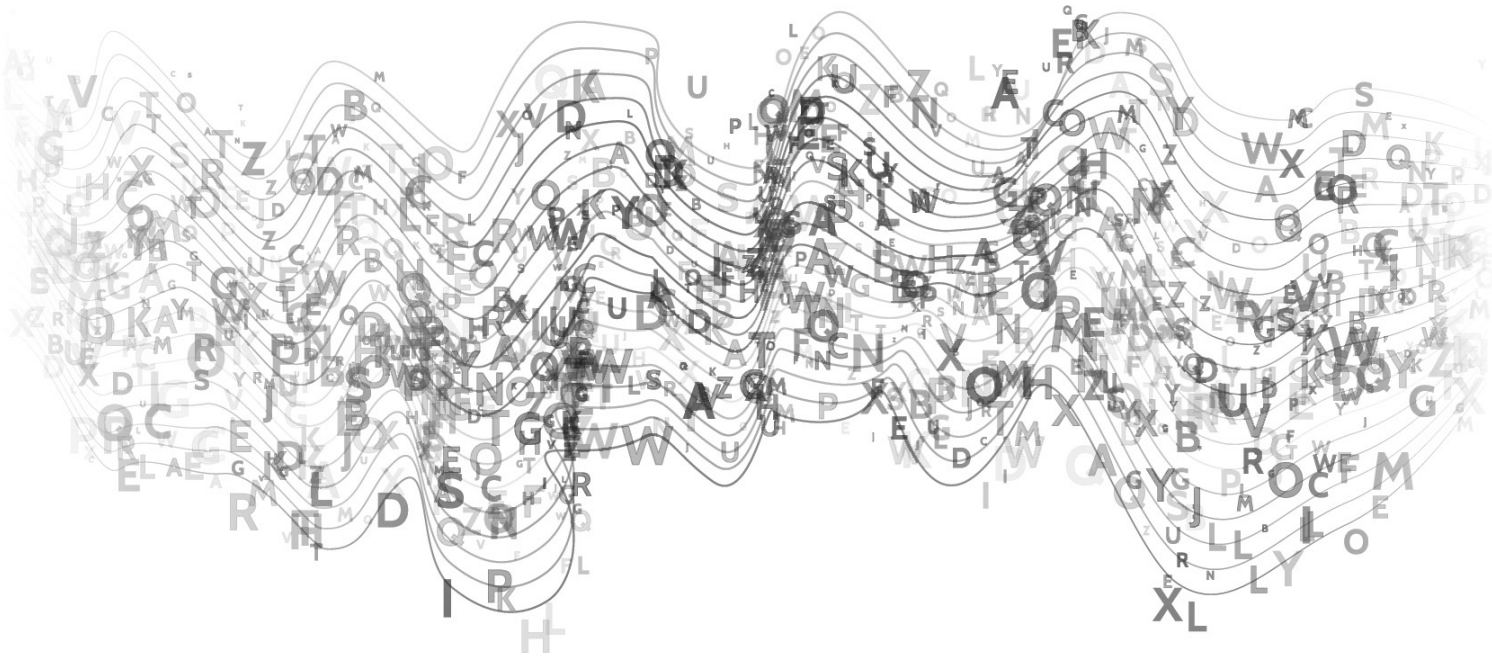
The primary audience for the guide includes IIS managers and staff with responsibility for ensuring IIS data quality. This may include individuals in various roles depending on the IIS program staffing structure and includes data quality specialists, data exchange staff, and/or interoperability or interface coordinators. Staff involved in the onboarding process and staff involved in the technical maintenance and development of IIS functionality may also benefit from the content shared in the guide. Immunization program staff in positions that interact with the IIS data may also find the guide of interest.

Methodology of Guide Development

To create the guide, AIRA assembled a workgroup of SMEs from the IIS community, Centers for Disease Control and Prevention (CDC) partners, public health consultants, and AIRA staff (see list of participants in the Acknowledgements section). During the initial phase of the project, existing IIS materials were gathered and reviewed to identify existing data quality guidelines for data at rest. The reviewed materials include AIRA's Modeling of Immunization Registry Operations Workgroup (MIROW) best practice guidelines on data quality, CDC Health Level 7 (HL7) Implementation Guides, CDC's IIS Functional Standards, and documents gathered directly from IIS programs. With support from a public health consultant and an AIRA project manager, the workgroup met via telephone from October 2017 through April 2018. The workgroup reviewed materials and developed recommendations for the guide. The consultant drafted and revised the guide based on input and feedback from the workgroup and others. Finally, the document was reviewed by AIRA staff, the AIRA board of directors, and the IIS community, with the final version completed in June 2018.

Primary resource materials reviewed for this topic include:

- AIRA Data Validation Guide for the IIS Onboarding Process, February 2017
- AIRA IIS Data Quality Practices: Monitoring and Evaluating Data Submissions, September 2017
- AIRA Immunization Information System (IIS) Implementation Guidance for a Shared Address Cleansing and Geocoding Service, May 2017
- AIRA Modeling of Immunization Registry Operations Workgroup (MIROW) Best Practices Guides:
  - Data Quality Assurance in Immunization Information Systems: Incoming Data, February 2008
  - Data Quality Assurance in Immunization Information Systems: Selected Aspects, May 2013
- HL7 Version 2.5.1 Implementation Guide for Immunization Messaging, v 1.5 and Addendum
- CDC Endorsed Patient Demographic and Vaccination Event Data Elements, version 4.0
- CDC IIS Functional Standards, version 4.0



SECTION II. DATA QUALITY DIMENSIONS AND INDICATORS

Data Quality Dimensions

In this section, we examine specific dimensions and indicators of IIS data quality. "A Data Quality Dimension is a recognized term used by data management professionals to describe a feature of data that can be measured or assessed against defined standards to determine the quality of data."[4]

A "feature" is a characteristic, attribute or facet of the data. The dimensions most commonly used to measure the quality of public health data are completeness, accuracy, and timeliness.[5,6] IIS are especially interested in these dimensions because they have ramifications for clinical decision making, vaccine tracking and accountability, vaccination coverage assessments, and public health research. The additional dimensions of consistency and validity, closely related to accuracy, are also relevant to IIS data quality assessment and will be explored in this section. The dimension of uniqueness is relevant to patient and vaccination deduplication and is briefly covered in Section III and Appendix C.

A Data Quality Dimension is a recognized term used by data management professionals to describe a feature (characteristic, attribute or facet) of data that can be measured or assessed against defined standards to determine the quality of data (DAMA UK Working Group on Data Quality Dimensions, 2013).

Dimension Definitions[7]

Completeness: The degree to which full information about a data set, record, or individual data element is captured in the IIS, i.e., the proportion of stored data measured against the potential of "100% complete."

Accuracy: The degree to which data correctly describes the "real-world" object or event being described.

Validity: The degree to which the data conforms to the syntax (format, type, range) of its definitions, i.e., to the rules of what is accepted or expected by the IIS.

Consistency: The absence of difference when comparing two or more representations of a thing against a definition.

Timeliness: The amount of time between the occurrence of the real-world event and its documentation in the IIS, i.e., the time lag between the date of vaccination or birth and the date the record is received by the IIS.

Uniqueness: No event, person, or data element is recorded more than once. Note: uniqueness is not covered as a separate dimension in this section but is addressed in the Section III discussion of special situations in Patient and Vaccination Record Uniqueness.

4 DAMA UK Working Group on Data Quality Dimensions. The Six Primary Dimensions for Data Quality Assessment Defining Data Quality Dimensions. 2013. p. 1. https://www.dqglobal.com/wp-content/uploads/2013/11/DAMA-UK-DQ-Dimensions-White-Paper-R37-1.pdf.
5 The 2008 MIROW Guide, Data Quality Assurance in Immunization Information Systems, discusses accuracy, completeness, and timeliness. See pp. 86-88. http://repository.immregistries.org/resource/data-quality-assurance-in-immunization-information-systems-incoming-data-1/.
6 Chen, H., Hailey, D., Wang, N., and Ping, Y. "A Review of Data Quality Assessment Methods for Public Health Information Systems," International Journal of Environmental Research and Public Health, 11(5): 5170-5207. 2014.
7 These definitions are derived from DAMA UK Working Group. "The Six Primary Dimensions for Data Quality Assessment."

Keeping the measures for each dimension separate from the other dimensions will aid in determining the root cause of data quality problems.

Completeness Indicators

Our definition of completeness is the degree to which full information about a data set, record, or individual data element is captured in the IIS. We assess this by measuring the proportion of stored data against the potential of "100% complete."[8]

> **Completeness:** The degree to which full information about a data set, record, or individual data element is captured in the IIS.

Completeness can be examined at the individual data element level, the individual patient or immunization record level, or the population level. Ensuring completeness of high-priority demographic and vaccination fields is often accomplished at the point of data entry into the IIS. Most IIS have a minimum/mandatory set of data items that need to be present for a record to be accepted into the IIS. For example, if an electronic data submission is missing primary demographic or vaccination information, then the message is often rejected during the HL7 processing. In the case of direct manual data entry, an IIS might not allow the record to be saved if certain key elements are missing. These early preventative actions by an IIS help ensure that incomplete records do not get into the system. Still, there may be methods for data to enter the IIS that do not have these controls. Older data might not have gone through the current process. Also, non-mandatory data elements might not have been reviewed in the incoming data phase yet are valuable to the IIS for IIS-specific program functions. Non-mandatory data elements may also help determine the validity of a record through data element cross-checks, as described in the Validity section. Population-level completeness, often referred to as IIS saturation levels, refers to the extent to which an IIS has demographic records for the full population and all the vaccinations they received. This is discussed more fully in the Population-Level Completeness subsection.

It should be noted that a high completeness level alone does not ensure a high quality of data. If a data item is mandatory, 100% completeness may be easily achieved, but accuracy can be lacking. Validity and consistency checks need to be performed to determine if the data items have been completed correctly. Knowing the accuracy of the data is essential in determining the proper significance of completeness rates. This is discussed in more detail in the sections below.

8 DAMA UK Working Group. "The Six Primary Dimensions for Data Quality Assessment."

Explanation of Completeness Tables

Table 1 and Table 2 provide completeness recommendations on demographic and vaccination elements, respectively, to examine data at rest. The priority assignment and target levels were determined by the Data at Rest Guide workgroup, with consideration of the target levels established in other documents. Many of the target levels are the same as those in the Phase Two guide.

> ***The list of data elements in Tables 1 and 2 is not all-inclusive. Each IIS should select and add data elements as indicated by its business needs.***

The priority assignment (high, medium, or low) for each element does not reflect the priority and importance of the element to the IIS overall. It reflects only the importance of the particular element to data quality analysis for data at rest. All of the elements listed in Table 1 and Table 2 can be found in the CDC *Endorsed Data Elements, Version 4* list.[9] However, not all of the Endorsed Data Elements are included in the tables below—only those deemed of most importance to data at rest assessment. IIS may want to evaluate additional data elements based on their own programmatic and business function needs.

The target-level column presents a range of completeness values for data elements, with the expectation that the individual IIS will establish a level within the given range specific to its inquiry. For example, an analysis that includes older data may have lower target-level expectations than those that look only at new data.

The source column lists outside resources that contributed to the selected target levels. Business Rules 104 and 105 from the *MIROW Data Quality Assurance in Immunization Information Systems: Selected Aspects* list the minimum/mandatory elements required for a data submission to be accepted.[10] Thus, these elements usually should be 100% complete. For more details on Business Rules 104 and 105, see the Glossary listing for "Minimum/Mandatory Data Elements." The source identified as "Workgroup" means that the SME Workgroup for the guide determined the recommended target levels in the absence of existing documentation.

9 See https://www.cdc.gov/vaccines/programs/iis/core-data-elements.html for more information about the Endorsed Data Elements.
10 http://repository.immregistries.org/resource/data-quality-assurance-in-immunization-information-systems-selected-aspects/.

Data Element Completeness Tables

Table 1 | *Patient Demographic Data Element Completeness Recommendations*

PRIORITY LEVEL	DATA ELEMENT	TARGET LEVEL	SOURCE OF TARGET LEVEL[11]	NOTES/CAVEATS	DQ-RELATED BUSINESS FUNCTION
High	Patient First Name	100%	MIROW- BR104/105 IISAR	• Minimum/mandatory element.	• Patient-Level Deduplication
High	Patient Middle Name or Initial	75-80%	Workgroup	• Target lower than first and last names because not everyone has a middle name. • Many providers do not send this, but it is very helpful for deduplication. • Previous guides do not include target levels for middle name.	• Patient-Level Deduplication
High	Patient Last Name	100%	MIROW- BR104/105 IISAR	• Minimum/mandatory element.	• Patient-Level Deduplication
High	Patient Date of Birth	100%	MIROW- BR104/105 IISAR	• Minimum/mandatory element.	• Patient-Level Deduplication • Data checks per Tables 3, 5
High	Patient Address: Street	85-95%	DQ Guides IISAR	• Patient address elements are not broken out in the two previous data quality guides. IISAR does break out these elements. • All address elements are important if doing mailed reminder-recall or geographic-based Vaccination Coverage Assessment. • Newer data should aim for the higher target level.	• Patient-Level Deduplication • Inactivation of patients who moved out of jurisdiction
High	Patient Address: City	85-95%	DQ Guides IISAR		
High	Patient Address: State	85-95%	DQ Guides IISAR		
High	Patient Address: ZIP	85-95%	DQ Guides IISAR		

11 Phase One and Phase Two Data Quality Guides informed the selection of target levels for this guide (and are abbreviated as DQ Guides). MIROW-BR refers to business rules from the AIRA MIROW Data Quality Assurance in Immunization Information Systems: Selected Aspects. BR 104/105 list the minimum/mandatory elements required for a data submission to be accepted, usually expected to be 100% complete. See Appendix B definition of minimum/mandatory for full list.

PRIORITY LEVEL	DATA ELEMENT	TARGET LEVEL	SOURCE OF TARGET LEVEL[11]	NOTES/CAVEATS	DQ-RELATED BUSINESS FUNCTION
High	Mother's First Name	90%	DQ Guides IISAR	• Applies only to submission of information about a minor. Responsible person fields may be substituted for mother's name. • Components of mother's name not broken out in other guides. • IISAR measures mother's first and last name separately • Count any responsible person's name for mother's name. • Important if doing reminder-recall. Middle name target is lower than other elements because not everyone has a middle name.	• Patient-Level Deduplication
High	Mother's Middle Name	75-80%	DQ Guides		
High	Mother's Last Name	90%	DQ Guides IISAR		
High	Mother's Maiden Name	90%	DQ Guides	• Applies only to submission of information about a minor.	• Patient-Level Deduplication
Medium	Patient ID (from data source)	100%	DQ Guides	• Refers to medical record number or other identifying number used by the submitting organization. • Useful for deduplication of patients within that organization.	• Patient-Level Deduplication • Data checks per Table 5
Medium	Patient Gender	95-100%	MIROW-BR104 DQ Guides IISAR	• Minimum/mandatory element. • Important if examining vaccination rates by gender.	• Patient-Level Deduplication • Data checks per Table 5, such as cross-checks of gender-specific vaccine

PRIORITY LEVEL	DATA ELEMENT	TARGET LEVEL	SOURCE OF TARGET LEVEL[11]	NOTES/CAVEATS	DQ-RELATED BUSINESS FUNCTION
Medium	Patient Race	95-100%	DQ Guides	• May be used to examine vaccination rates by race and ethnicity.	• Data checks per Table 5
Medium	Patient Ethnicity	95-100%	DQ Guides		
Medium	Patient Phone	90-95%	DQ Guides	• Important if doing reminder-recall by phone call or text message. • For minors, the responsible party's phone number is often used.	• Patient-Level Deduplication
Medium	Patient Phone Type	90%		• Current gap in ability to define phone type being sent by EHR. • If using IIS to email or text reminder-recalls, these elements will be important for targeted age groups. If not, IIS may want to set lower target levels. • For minors, the responsible party's phone type, email address, and language is often used.	• Patient-Level Deduplication
Low	Patient Email Address	90%			
Low	Patient Primary Language	90%			

Table 2 | *Vaccine Data Element Completeness Recommendations*

PRIORITY LEVEL	DATA ELEMENT	TARGET LEVEL	SOURCE OF TARGET LEVEL[12]	NOTES/CAVEATS	DQ-RELATED BUSINESS FUNCTION
High	Vaccine Administration Date	100%	MIROW- BR104/105 IISAR	• Minimum/mandatory element. • Required for dose decrementing and inventory management and vaccine accountability for VFC-eligible children. • Necessary for accurate forecasting (i.e., recommendations of next vaccine due). • Required for Vaccination Coverage Assessment.	• Vaccine Deduplication • See Data Checks per Tables 3, 4, 5
High	Vaccine Product Type	100%	MIROW-BR105	• Minimum/mandatory element. • Required for dose decrementing and inventory management and vaccine accountability for VFC-eligible children. • Required for Vaccination Coverage Assessment.	• Vaccine Deduplication • See Data checks per Tables 3, 5, 6
High	Vaccination Event Record Type (admin/ historical)	100%	MIROW-BR105 IISAR	• Minimum/mandatory element. • Aka "Vaccine Event Information Source" per HL7 Implementation Guides.	• Vaccine Deduplication • See Data checks per Tables 4, 5, 6
High	Vaccine Manufacturer	100% of administered vaccines	MIROW-BR116 IISAR	• For administered vaccines only.	• Vaccine Deduplication • See Data checks per Table 3

12 Phase One and Phase Two Data Quality Guides informed the selection of target levels for this guide (and are abbreviated as DQ Guides). MIROW-BR refers to business rules from the AIRA MIROW Data Quality Assurance in Immunization Information Systems: Selected Aspects. BR 104/105 list the minimum/mandatory elements required for a data submission to be accepted, usually expected to be 100% complete. See Appendix B definition of minimum/mandatory for full list.

PRIORITY LEVEL	DATA ELEMENT	TARGET LEVEL	SOURCE OF TARGET LEVEL[12]	NOTES/CAVEATS	DQ-RELATED BUSINESS FUNCTION
High	Vaccine Lot Number	100% of administered vaccines	MIROW-BR105 IISAR	• For administered vaccines only. • Required for dose decrementing and inventory management and vaccine accountability for VFC-eligible children. • IIS-specific requirements for tracking lot numbers of private source vaccine (and target level) may vary. • Important for identifying patients affected by vaccine recall.	• Vaccine Deduplication • See Data checks per Tables 4, 5
High	Vaccine Expiration Date	90-100% of administered vaccines	MIROW-BR118	• For administered vaccines only. • Required for dose decrementing and inventory management and vaccine accountability for VFC-eligible children.	• Vaccine Deduplication • See data checks per Table 4
High	Vaccine Eligibility at Dose Level	100% of administered vaccines among VFC providers	MIROW-BR122 DQ Guides IISAR	• For administered vaccines only. • Required for dose decrementing and inventory management and vaccine accountability for VFC-eligible children. • Some IIS also request submission of funding source—aka dose-level public/private indicator—to help with dose-decrementing from inventory and for vaccine accountability purposes.	• See data checks per Table 4 • N/A for non-VFC providers and for adult patients, unless IIS has need to track vaccine funding source for certain adult vaccines

PRIORITY LEVEL	DATA ELEMENT	TARGET LEVEL	SOURCE OF TARGET LEVEL[12]	NOTES/CAVEATS	DQ-RELATED BUSINESS FUNCTION
Medium	Vaccine Dose Volume and Unit	90% of administered vaccines	DQ Guides	• For administered vaccines only.	• Vaccine Deduplication
Medium	Vaccine Site of Administration	90% of administered vaccines	MIROW-BR119	• For administered vaccines only.	• Vaccine Deduplication • See data checks per Tables 3, 5
Medium	Vaccine Route of Administration	90% of administered vaccines	MIROW-BR119	• For administered vaccines only.	• Vaccine Deduplication • See data checks per Tables 3, 5
Medium	Vaccine Administering Provider: Name, Suffix	90% of administered vaccines	DQ Guides	• For administered vaccines only.	• Needed data quality checks to identify patterns by provider
Low	Vaccine Information Statement (VIS Information: Type, Publication Date, Date Given to Patient	90% of administered vaccines		• Many IIS do not process or store this data. • Applies to administered vaccines only. • Expected to be captured in the IIS if the IIS is used as the primary vaccination event record (e.g., mass vaccination clinic).	• Documentation required for compliance with Vaccines for Children program, but not required to be in IIS if providers maintain in their own system

Population-Level Completeness

Population-level completeness is affected by IIS provider participation levels, with higher participation usually meaning more people and their vaccinations will be included in the IIS. However, even with full provider participation, records may be completely or partially missing. Thus, it is important to evaluate how well the IIS population matches the jurisdiction population. An approximation of completeness at this level can be derived by comparing numbers in the IIS to sources such as the US Census Bureau. This can be broken out by age group or geographic area or other demographic characteristics. Resources for external population numbers are described in AIRA's *Analytic Guide for Assessing Vaccination Coverage Using an IIS*.[13] Completeness of vaccination records can also be evaluated by comparing IIS-generated vaccination coverage results to other sources, such as the National Immunization Survey (NIS).[14] Differences in findings can be due to a variety of factors, including under-population of the IIS, incomplete vaccination histories, duplicate records, and active records for people who have moved out of the area. The AIRA guide *Comparing and Communicating Vaccination Coverage Estimates from Immunization Information Systems, the National Immunization Survey, and Related Assessments* provides instructions on how to compare IIS coverage results to the NIS and other sources.[15] The consistency-related measures in Table 6 can also assist with evaluation of overall data completeness.

Accuracy Indicators

Our definition of accuracy is the degree to which data correctly describes the "real-world" object or event being described. Accurate data means that records in the IIS exactly mirror the person or event. For example, the data describing a vaccination

Accuracy: The degree to which data correctly describes the "real-world" object or event being described.

event should match exactly with what happened in the clinical encounter, whether or not it is clinically appropriate. For example, administering a vaccine outside of the routine schedule may have actually occurred in the clinic, or it may reflect an error in the data (e.g., wrong vaccine code submitted, incorrect date of administration, incorrect birth date). Best practice dictates that the data reflect the actual event even if outside of standard practice. Differentiating between technical issues and clinical practice issues can require close analysis. Comparing IIS data to the primary source data—usually the medical record—is the gold standard for evaluating accuracy. However, it is often impractical due to the IIS resources required. Asking providers to periodically review and compare a random sample of their own submissions to the IIS record is an effective method to try to share the workload. This can also increase provider ownership of data quality. Using IIS or immunization program staff to compare IIS records to the clinical record may be best saved for serious or frequent data issues that can be resolved only by using this method. Fortunately, IIS can implement alternative time-efficient strategies for assessing accuracy and determining the necessity of going to the clinical record. For example, automated reports can be developed that look for internal data inconsistencies suggesting problems with accuracy. While the results may indicate a need to follow up with the source data, use of internal measures related to validity and consistency can help narrow the search for accuracy issues. The validity and consistency indicators below offer specifics on what to look for.

13 See http://repository.immregistries.org/resource/data-quality-assurance-in-immunization-information-systems-incoming-data-1/, pp 13-17.
14 See https://www.cdc.gov/vaccines/imz-managers/nis/index.html.
15 See http://repository.immregistries.org/resource/comparing-and-communicating-vaccination-coverage-estimates-from-iis-nis-and-related-assessments/.

Validity Indicators

Our definition of validity is the degree to which the data conform to the syntax (format, type, range) of its definition—i.e., to the rules of what is accepted or expected by the IIS. Validity applies at the data item level and record level (for combinations of valid values) and may be applied at the aggregate record level (over a large volume of records) in certain circumstances. Database rules should provide the allowable types (string, integer), the format (length, number of digits), and range (minimum, maximum, or contained within a set of allowable values).[16] When data validity is applied at the single data item level, it can be thought of as "within single field" data validation. When applied at the record level, it can be thought of as "cross-field" data validation. At the aggregate level, data validity can be thought of as trend analysis.

Some cross-field validity issues can be discovered by a comparison of related data fields within a record in a search for contradictory items. An example is comparing the vaccine type to its accompanying manufacturer within a vaccination event record. If a query finds varicella vaccine doses with a manufacturer of Medimmune, then this reflects a data error since Medimmune does not manufacture varicella vaccine. Either the vaccine type or the manufacturer is incorrect. Other validity issues may surface by comparing IIS data elements to tables of expected and allowed values. An example of this is comparing a vaccine type to a list of vaccines currently available in the US. The vaccine DTP (diphtheria/tetanus/pertussis) for example should not show up as a currently administered vaccine. A one-time incidence of a contradiction may not be cause for concern. However, when patterns or high volumes of particular errors occur, follow-up is needed to discover the cause. The validity indicators in Table 3 below describe issues that can be uncovered by cross-validation processes.

Validity: The degree to which the data conform to the syntax (format, type, range) of its definition.

16 DAMA UK Working Group on Data Quality Dimensions.

Table 3 | *Indicators of Validity Issues – Cross-Field Contradictions (Unrelated to ACIP Recommendations)*

PRIORITY LEVEL	INDICATOR[17]	NOTES
High	Vaccine administered is not yet available to clinicians.	• Example: A new flu vaccine becomes available on 07/31/2017; however, a record is received indicating administration on 06/02/2017.[18]
High	Lot numbers that violate manufacturer lot number patterns.	• Will need to cross-validate lot number compared to known manufacturer lot number patterns per footnoted reference.[19] • Example: Unit of Use lot number pattern documented in the IIS does not follow that specific manufacturer's lot number pattern.
High	Lot numbers that violate format, syntax, or range expectations based on provider inventory.	• Example: Where available, lot numbers could be compared with inventory of VFC or other state- supplied vaccine for lot number confirmation.
High	Vaccination administration date is after the vaccine expiration date for the corresponding vaccine lot.	• Discovered through cross-validation of administration date and expiration date of associated lot number.
Medium	Vaccination date is before patient date of birth.	• May be flagged in HL7 processing so may not show up in data at rest—may be N/A. • Or historical/transitional data entry might cause this.
Medium	Vaccination date is after the submission date (i.e., vaccination date is in the future).	• May be flagged in HL7 processing so may not show up in data at rest—may be N/A.
Medium	Manufacturer and vaccine product contradict one another.	• Cross-validations can identify inconsistencies in manufacturers and vaccine products, which can indicate a coding or entry problem—especially important to DQ if it shows wrong vaccine being administered. NDC also can be cross-validated against both manufacturer and product code.
Medium	Submitted vaccine descriptions and codes contradict one another.	• Example: CVX code 144 (intradermal flu vaccine code) and the vaccine name Pediarix® submitted for one immunization event.
Medium	Vaccine administration route contradicts the administration site.	• Example: A vaccination administered orally cannot be administered with a site of left thigh or right arm. The site or route code is wrong, maybe due to a coding issue.

17 Some IIS have business rules that prevent these scenarios from being processed and stored. Each IIS can assess and determine if implementation of such business rules is warranted.
18 CDC's Epidemiology and Prevention of Vaccine-Preventable Diseases (The Pink Book) contains a list of current US vaccines, along with the dates of approval in its Appendix B.
19 AIRA. Vaccine Lot Number Patterns: Unit of Sale/Unit of Use Guidance. June 7, 2017. http://repository.immregistries.org/resource/guidance-on-unit-of-sale-unit-of-use-lot-numbers/.

PRIORITY LEVEL	INDICATOR[17]	NOTES
Medium	Vaccine reported as administered in US has never been available in US.	• Validate against a comprehensive list or table of vaccines ever administered in US.[20]
Medium	Vaccine administered is not a vaccine that was ever available and is not in the pipeline of new vaccines.	
Medium	Administered vaccinations submitted with CVX codes indicating unspecified formulation for a vaccine.	• Unspecified codes should be used only for historical vaccines where precise formulation is unknown. If used with administered vaccine, might indicate that administered/historical code is incorrect or that provider submitted an incorrect vaccine code. • Examples: ◦ CVX code 107 – "DTaP, unspecified formulation" ◦ CVX code 17 – Haemophilus influenzae type b vaccine, conjugate unspecified formulation
Medium	Vaccinations with CVX codes indicating unspecified formulation for a vaccine that also have lot number or NDC entry.	• A provider should not be able to submit a lot number or NDC if he does not know the specific type of vaccine the patient received.
Medium	Submission of "unknown" for various fields for an administered immunization (e.g., manufacturer).*	• A submission of "unknown" for a particular value may appear to be a complete submission to the IIS, but it is not precise and it can have implications for data quality checks and data use.
Low	Birth date is after the submission date (i.e., birth date is in the future).	• May be flagged in HL7 processing so may not show up in data at rest.
Low	Vaccine records with mismatched vaccines and dose sizes.	• Vaccines are available only with certain dose sizes. A mismatch between the expected and observed vaccine dose size may mean either an incorrect clinical practice was properly captured or that a correct clinical practice was improperly captured.
Low	Individual address components that contradict each other.	• Examples: ◦ Zip code and state mismatch ◦ Zip code and city mismatch

**These may also be useful consistency indicators if comparing the proportions to a previous period of time.*

20 See CDC's *Epidemiology and Prevention of Vaccine-Preventable Diseases* (The Pink Book). It contains in its Appendix B a list of discontinued US vaccines and, for some of them, the date ranges when they were available.

Cross-field validity issues that are related to deviations from the Advisory Committee on Immunization Practices (ACIP) recommendations are listed in Table 4.

Table 4 | *Indicators of Validity Issues – Cross-Field Contradictions (Related to ACIP Recommendations)*

PRIORITY LEVEL	INDICATOR (ALL APPLY TO ADMINISTERED IMMUNIZATIONS)	NOTES
High	Vaccination administration date is after the vaccine expiration date for the corresponding vaccine lot.	• Discovered through cross-validation of administration date and expiration date of associated lot number.
High	Vaccination other than hepatitis B at birth and/or before 1 month of age.	• Indicates coding or data entry problem. Possible but rare actual clinical practice. • Discovered through cross-validation of DOB and date of vaccine administration.
High	Vaccination before the minimum patient age or after the maximum patient age for a particular vaccine group or product.	• Example: Rotavirus vaccine administered after 8 months, 0 days of age. • Discovered through cross-validation of DOB (to derive age) and date of vaccination.
High	Vaccine funding source and patient VFC eligibility contradict one another.	• Example: Private vaccine given to VFC-eligible child; state-supplied vaccine given to non-VFC eligible individual. • Discovered through cross-validation of VFC funding source value and patient VFC eligibility.
High	Patients with an unexpected total number of immunizations for their age.	• Examples: ◦ 20+ immunizations before age 6 months ◦ 30+ immunizations before age 2 years • Discovered through cross-validation of patient DOB (with derived age) and total number of vaccinations recorded in the IIS. Likely results from issues with IIS patient or vaccine deduplication; however, over-vaccination can occur, though unlikely if volume is set high enough as in examples above.

PRIORITY LEVEL	INDICATOR (ALL APPLY TO ADMINISTERED IMMUNIZATIONS)	NOTES
Medium	Proportions of vaccine types violate expectations.	• Using the ACIP recommended schedule, can look for vaccine types that should have similar numbers. For example: ◦ Numbers of DTaP and PCV doses should be similar for children under 2 years of age. ◦ Number of MMR doses should be significantly less than number of DTaP doses. ◦ Unexpected volume of unusual products, such as immune globulin (IG) at higher numbers than usual. • Unexpected differences may indicate issues with how vaccines are mapped from the EHR to the electronic interface.
Medium	Vaccine reported as administered was at one point available but was no longer in distribution, or vaccination date is outside of the US licensure date range for the product.	• May be submitted as historical if date is appropriate but should not be reported as administered.
Medium	Vaccine route and/or site contradictory for given patient's age.	• Example: DTaP administered to 15-month-old in left deltoid. Discovered through cross-validation of site of administration and DOB (age of child on vaccination date). • Caution: Monitor volume changes. Some differences may be expected due to special circumstances, such as baby with hip dysplasia cast may receive at alternate site but rare.

Validity issues related to syntax, format, and range within single fields are listed in Table 5.

Table 5 | *Indicators of Validity Issues Within Single Fields*

PRIORITY LEVEL	INDICATOR (ALL APPLY TO ADMINISTERED IMMUNIZATIONS)	NOTES
High	Submission of invalid client demographic data.	• Note that these cases may be flagged or rejected in HL7 processing. • Examples: ◦ Name: submission of generic name such as "Baby," "Mickey Mouse," "Test," etc. ◦ Text field (name, street, city, etc.): contains symbols that are not part of conventional text fields, such as $%"'-+=~`#!@*^, as these would likely hamper deduplication efforts. • Medical record number: follows Social Security number format. • Email: does not contain "@"; does not contain a period.
High	High volume of administered immunizations or dates of birth with dates of 01/01/YYYY or MM/01/YYYY.*	• The first month of the year and/or the first day of the month may have been used as a stand-in date when the precise immunization date is not known. Historically, hand-written or paper immunization records often contained only month and year for vaccination date. In addition, in certain populations with unknown birth dates, a default date of January 1 may have been used in their documentation.
High	Submission of placeholder data for numeric fields such as phone number, patient ID values, lot number, etc.	• May be flagged or rejected in HL7 processing. • Repeated or consecutive numbers are often indicators of placeholder numeric data. • Strings of extraneous symbols (#$%@^*!) may indicate placeholder data.
High	Lot numbers that violate format, syntax, or range expectations.[21]	• Differs from lot number measures in Table 3, as this indicator is single-field evaluation for known issues with lot numbers. • Examples: ◦ Numbers that start or end with certain combinations of characters, such as MED, SKB, LOT, PENT, DTAP, etc. ◦ Lot numbers should be represented only by combinations of letter(s), number(s), and/or dash(es).

21 AIRA. Lot Number Validation Best Practices. June 2015. http://repository.immregistries.org/resource/mirow-micro-guide-lot-number-validation-best-practices/.

PRIORITY LEVEL	INDICATOR (ALL APPLY TO ADMINISTERED IMMUNIZATIONS)	NOTES
Medium	Proportions of expected values for a given demographic field (e.g., race, ethnicity, gender, patient status, etc.) violate expectations.*	• Examples: ◦ For gender field: number of records for male comprise 80% of total compared to 20% for female. ◦ For race field: 100% of records for a given period of time have same race selected. ◦ For patient status field: a high (and unexpected) proportion of records show inactive status. • May indicate a default value being assigned to the data element or incorrect coding. • May also be captured in consistency measures.
High	Any vaccine information field with invalid syntax, format, length, or range.	• Extraneous symbols (e.g., $#%&*~) challenge automated data processes such as inventory decrementing or deduplication (arguably an accuracy measure, not validity). • Maximum field length indicates a risk the data was truncated (e.g., 40 characters out of a 40-character limit). • Lot number may be expected to have format of letter, letter, hyphen, number, number, number, number, letter but in fact has an alternate syntax. • Number expected (dose size), but field contains letters, logical values (TRUE, FALSE), etc.

**These may also be useful consistency indicators if comparing the proportions to a previous period of time.*

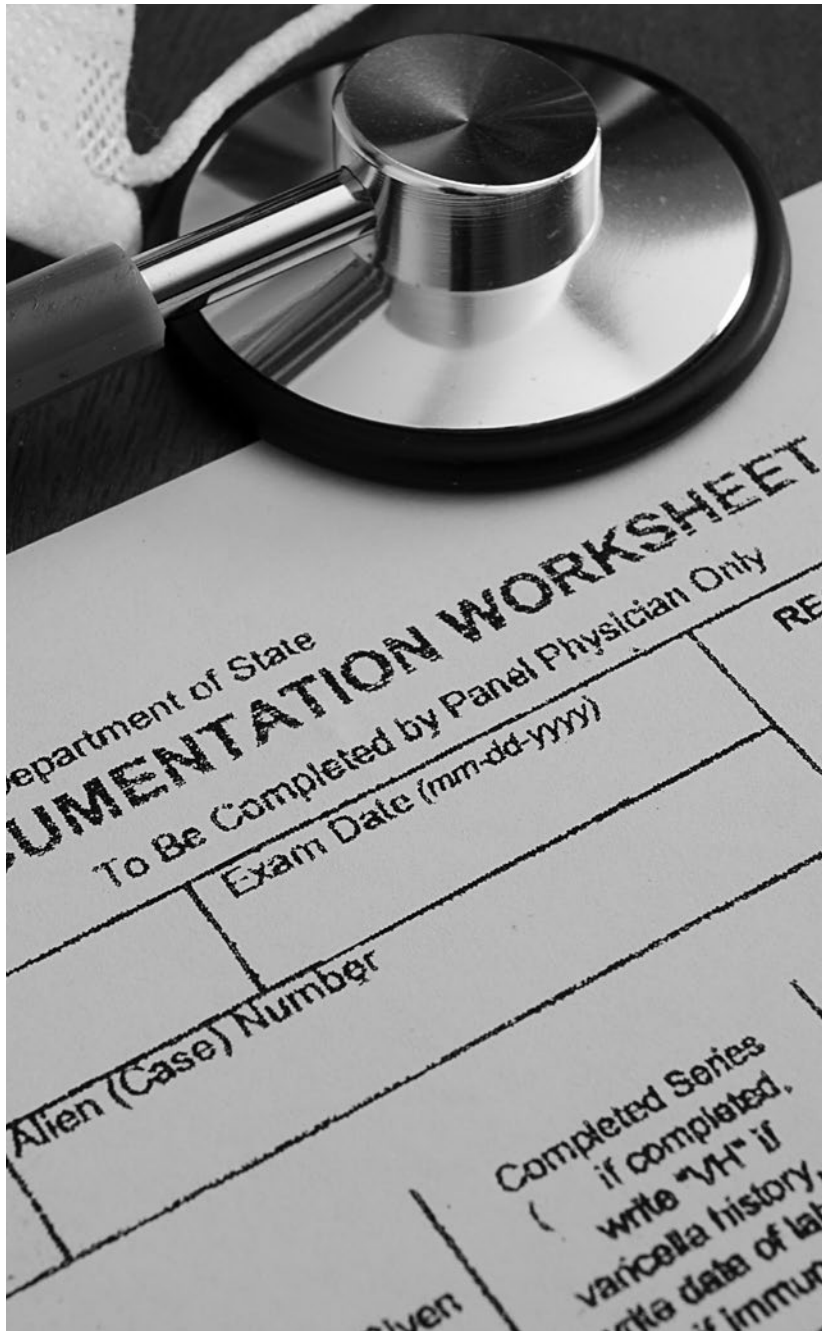
The concept of validity also can be applied to the types of vaccines received and whether or not the vaccine types are appropriate for the patient's demographics (e.g., age, interval between doses, and gender). The IIS expects to receive vaccine codes that are recommended for the patient's age and previous vaccination history according to ACIP. Deviations from the ACIP recommendations may indicate clinical practice validity issues (but not necessarily data validity issues). They may ***accurately*** represent the vaccination encounter when vaccine administration errors occur and doses are administered outside of the ACIP schedule. They may also represent doses that were given according to the ACIP schedule at the time of administration, noting that the ACIP recommendations change over time. On the other hand, they may represent data entry or other technical issues. Teasing out the difference between clinical practice and technical errors can be difficult. For example, if zoster vaccines show up for many one-year-olds, we might suspect a documentation (or coding) error since zoster is normally recommended for adults over 50, not children. On the other hand, MMR presenting with a pattern of administration at 11 ½ months of age, when the minimum age is 12 months, may reflect a provider's practice of administering this vaccine two weeks early. The guide does not intend to address correcting practice errors, which should be caught and addressed during incoming data review.[22] Rather, the guide will focus on analyzing trends that indicate data issues.

22 See http://repository.immregistries.org/resource/iis-data-quality-practices-monitoring-and-evaluating-data-submissions/.

Consistency Indicators

Our definition of consistency is the absence of difference when comparing two or more representations of a thing against a definition. Data consistency is assessed by measuring a data item against itself or its counterpart in another data set or database. This can be done by analyzing patterns and/or value frequency.[23] For example, the IIS can produce a breakdown of vaccine types administered last month and compare it to vaccine types administered during the same month last year. This may reveal inconsistencies in proportions of specific vaccines administered, reflecting possible data quality issues that need further exploration. Comparing external data sources to the IIS records can also indicate potential data quality problems. For example, the number of combination vaccines recorded in the IIS for a given period of time can be compared to the known state supply during that time, with significant differences indicative of possible issues. Table 6 presents indicators of potentially inconsistent reporting that can be analyzed through consistency-related measures.

Consistency: The absence of difference when comparing two or more representations of a thing against a definition.



23 DAMA UK Working Group on Data Quality Dimensions.

Table 6 | *Indicators of Potentially Inconsistent Reporting by Health Organizations*

PRIORITY LEVEL	INDICATORS OF POTENTIALLY INCOMPLETE REPORTING	NOTES
High	The volume of immunizations for a specific period is not consistent with program expectations, previously noted patterns, or vaccine ordering.	For example, total number of submitted immunizations may show a significant unexpected change from a previous time period. Or they may show a change in the number and/or proportion of certain vaccines being administered or numbers not matching vaccine ordering patterns (for VFC). Or they may show an unusually high proportion of historical vaccines. This is best analyzed by examining trends over time. Will likely require further investigation at the provider level if no cause is obvious.
High	Patient age breakdown in the IIS is not as expected or shows an unexplained change in proportions.	This measure analyzes whether all age groups are represented in the IIS at the levels expected (e.g., compared to US Census or other external data sources). Could be analyzed at a county/jurisdiction level, by age cohort or full database, as well as at provider level.
Medium	Number of common combination vaccines is lower than expected.	Expectation of number may be based on known state supply and/or vaccine availability or on prior history of numbers in IIS. Differences may indicate issues with how vaccines are mapped from the EHR to the electronic interface—e.g., combination vaccines may be mapped to their individual antigen components.
Medium	Historical immunizations are not represented in a proportion consistent with program expectations or previously noted patterns.	May indicate that historical immunizations are being left out of data submissions to the IIS or are not being flagged appropriately as historical. Or over time, a provider may simply have fewer historical records to submit, especially if a stable patient population. Analysis could examine correlation with EHR software to see if issues are related to particular EHR software.
Low	IIS-generated vaccination coverage assessment results are much higher or lower than expected when comparing to previous IIS-generated assessments or to other relevant sources.	Differences in results can result from a variety of factors, including IIS under-population, incomplete vaccination histories, duplicate records, and active records for people who have moved out of the area. In addition, assessment methodologies can differ and affect results. For more information on comparing IIS vaccination coverage estimates to other sources, see AIRA's *Comparing and Communicating Vaccination Coverage Estimates from Immunization Information Systems, the National Immunization Survey, and Related Assessments.*[24]

24 http://repository.immregistries.org/resource/comparing-and-communicating-vaccination-coverage-estimates-from-iis-nis-and-related-assessments/.

PRIORITY LEVEL	INDICATORS OF POTENTIALLY INCOMPLETE REPORTING	NOTES
Low	Change in proportion of minimum interval violations or in proportion of age violations (too young or too old).	Changes over time may indicate a change in actual practice, perhaps with new staff less familiar with the nuances of the schedule causing an increase in interval violations. Or changes may reflect systemic data entry problems (wrong vaccine codes selected, e.g. for live virus vaccine).
Low	Shifts in average proportion of records (patients or vaccinations) on the manual deduplication queue are greater than expected.	An increase in average record proportion (e.g., to 4% from 2% of all received shots) may indicate a decrease in data quality, as patient or shot pairs are harder to deduplicate. Conversely, if the shift is a decrease (e.g., 4% to 2% of all received shots) in record proportion, it may indicate an increase in data quality.

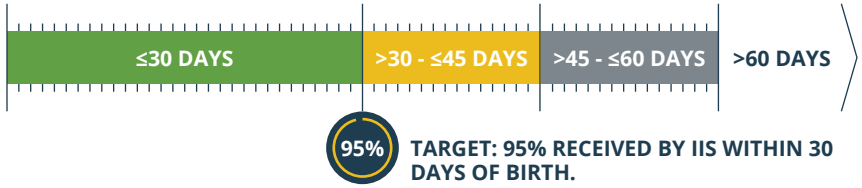
Timeliness Indicators

We define timeliness as the amount of time between the occurrence of the real-world event and its documentation in the IIS. For example, the time lag between the date of vaccination or birth and the date the record is documented in the IIS is a measure of timeliness. Timeliness is perhaps easiest and most important to assess during the ongoing review process at the provider level. Examining data at rest for timeliness is important in order to understand how well and reliably the IIS is able to provide data for various intended uses. For example, to be useful for clinical decision support, immunization data needs to appear in the IIS soon after the event occurs; timely data may help avoid over-vaccination. Similarly, if the IIS is used for vaccine tracking and accountability, records need to be updated frequently. In the event of an outbreak, timely data may help to assess vulnerable patients at the population level. In general, if data is not entered into the IIS in a timely manner, then none of these functions can be accomplished appropriately. The IISAR includes a question about vaccination record timeliness. The categories are ≤3 days, >3−≤14 days, >14−≤30 days, and >30 days. The target is to receive 90% of administered doses within three days. For newborns, the IISAR timeliness categories for receipt of birth records are ≤30 days, >30−≤45, >45−≤60, >60. The target is 95% within 30 days of birth. Some IIS have policies or laws requiring (or encouraging) submission within two to four weeks of vaccine administration. Policies may have been developed prior to the existence of real-time EHR interfaces. Now that this type of interface is becoming the norm, some IIS expect data to be submitted much more quickly—with targets ranging from 24 hours or less to one week. Currently, there is a great deal of variation in timeliness expectations among IIS. As the IIS community moves forward with implementing standardized data quality practices, it is recommended that more specific timeliness targets be developed.

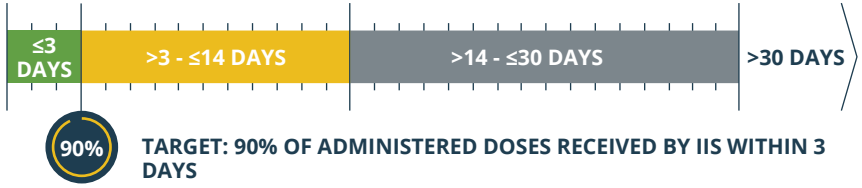
Timeliness: The amount of time between the occurrence of the real-world event and its documentation in the IIS.

Timeliness for vaccination record submission is best monitored continuously at the provider level as part of the ongoing review of incoming data. The Phase Two guide describes timeliness considerations in more detail. For data at rest, monitoring trends in timeliness may be of most interest. Comparing timeliness by variable, such as type of provider, data submission type, and EHR platform, may provide interesting information that can lead to interventions that identify and solve bottlenecks and improve timeliness performance.

Birth Record Timeliness Measures



Vaccination Timeliness Measures



SECTION III. SYSTEMIC DATA QUALITY CHALLENGES AND STRATEGIES

Keeping data clean, up-to-date, and standardized is a challenge for IIS. Algorithms that correctly match, link, and purge records are challenging to develop. This section of the guide offers high-level suggestions for processes that can help with keeping data clean and accurate. This includes address cleansing, address correction, changing the status of demographic records that should be inactive, fixing bad merges, and dealing with duplicate records at both the patient and vaccine levels.

Systemic data challenges:

- Keeping addresses standardized and up to date
- Inactivating records of patients who are deceased or have moved out of jurisdiction
- Maintaining separate records for multiple births (e.g., twins)
- Preventing duplicate records when birth data is submitted by both birthing facilities and Vital Statistics

Patient Address Accuracy: Improvement Through Address Cleansing and Correction

According to the United States Census Bureau, 11.2% of the US population age one year and older changed residence between 2015 and 2016.[25] This represents a staggering number of individuals with a change in address each year. Of the 35 million individuals who move each year, up to one third do not update their addresses with the United States Postal Service (USPS).[26] In 2009, researchers at the University of Michigan generated reminder-recall notices from the Michigan Community Immunization Registry (MCIR) for children 6 months through 19 years of age. They found that 26% of the 20,000+ mailed reminder-recall notices were returned as "undeliverable."[27] For IIS, all this means that "good" address data can degrade quickly due to address changes and other small differences in how address data is recorded. Incorrect addresses can affect IIS data quality in many ways. Negative impacts include the creation of duplicate records and inflation of the IIS denominator, decreased accuracy of geographic-based vaccination coverage reports, and returned/undeliverable reminder-recall mailings. To improve address data quality, some IIS take advantage of services that provide address cleansing and geocoding.

25 https://www.census.gov/newsroom/blogs/random-samplings/2017/01/mover-rate.html.
26 Health Management Technology website http://www.healthmgttech.com/healthy-data-helps-patients-and-providers-flourish.php.
27 Dombkowski, K.J. et al. Assessing the burden of undeliverable immunization reminder and recall notifications. 2011.

Address Cleansing

Address cleansing is the process for standardizing and correcting addresses within the IIS. This process does not match individuals with addresses but, rather, ensures that the proper address conventions are followed and that a physical mailing address actually exists. Postal address verification is a part of the process and is used to check the validity and deliverability of a physical mailing address. The process is also known as address standardization, address validation, address verification, and Coding Accuracy Support System (CASS) certification.[28] Before an address can be certified by the USPS as deliverable, it must first be standardized—with spelling, abbreviations, capitalization, and formatting all conforming to USPS rules.

> Standardization of addresses allows addresses to be more easily compared and improves the quality of IIS reporting capabilities.

The USPS offers address verification services directly to businesses, and it licenses its services to third-party companies that provide CASS certification for mailing lists.[29] These third-party vendors typically provide batch processing of address lists in CSV or Excel format. A CASS-certified service standardizes a mailing list, updates outdated addresses (but not at the person level), verifies that addresses are valid and complete, and provides geocoding.[30] CASS-certification also enables an organization to get the lowest rates for bulk mailing.

Initially, on start-up of a cleansing service, the IIS should run all existing records through the process. Then, as newly introduced records or updates for existing records are submitted, the IIS can push them through a routine, automated cleansing process.

According to the *IIS Implementation Guidance for a Shared Address Cleansing & Geocoding Service*, published by AIRA in 2017, the benefits of address cleansing and geocoding include:

- All addresses in the IIS database can be verified and formatted in accordance with the USPS database and standards.
- Standardization of addresses can improve patient-level deduplication by eliminating variability in addresses evaluated by the IIS match algorithm.
- Address cleansing and standardization of all address elements improve the quality and accuracy of reports run by various address and geographic parameters (e.g., Assessment, Feedback, Incentives, and eXchange (AFIX) and other coverage assessments).
- Improved address quality may help to reconcile the denominator of patients in the IIS with the actual census population by improving the deduplication process.
- Geocoding improves mapping capabilities, allows programs to look at data in different ways and enables better target intervention.
- For reminder-recall, address cleansing can improve mailing success by avoiding mailings to invalid addresses, decreasing the cost of unnecessary postage, allowing for discounted USPS bulk mailing rates when CASS-certified, and contributing to the end-goal: increasing patient response rates to mailed reminder-recall efforts.
- Integration of the address service into the IIS ensures that a verified/standardized address is not overwritten by a "bad" address and ensures that the address syntax (city, state, zip, county, and latitude/longitude) remain intact at all times. This can have an added benefit of prohibiting users from artificially

28 CASS is a term and system used by the United States Postal Service.
29 https://postalpro.usps.com/address-quality#cat-subsection-0.
30 One CASS-certified address cleansing service, SmartyStreets, is under contract with AIRA to assist IIS in updating their address data. https://smartystreets.com. See AIRA's IIS Implementation Guidance for a Shared Address Cleansing & Geocoding Service. http://repository.immregistries.org/resource/iis-implementation-guidance-for-a-shared-address-cleansing-and-geocoding-service/.

changing address elements to improve patient counts or coverage assessment outcomes.

- Standardization of addresses creates an opportunity to leverage or improve household grouping functions by identifying all family members associated with a particular address.

The corrections obtained from an address cleansing service can range from minor to critical. For example, the corrections may be as small as changing the nomenclature from "Street" to "St" or changing apartment designation from "#" to "Apt." More significant is the ability to correct a zip code that is simply wrong for the street, city, and state on record. Even more important is the ability to determine if a given address is invalid and not deliverable (i.e., does not exist). Table 7 offers examples of corrections received from an actual web-based correction service.[31]

Table 7 | *Examples of Addresses Before and After Cleansing*

SUBMITTED BY IIS	CORRECTED BY ADDRESS CLEANSING PROVIDER
Example 1: 78-7070 Alii Dr. #D-101 Kailua-Kona, HI 96740	78-7070 Alii Dr Apt D101 Kailua Kona HI 96740-4515 • Found 1 valid address • Matched street and city and state • Confirmed entire address • Fixed city/state spelling • Fixed abbreviations
Example 2: 1009 NE 71st Street Seattle, WA 98136	1009 NE 71st St Seattle WA 98115-5636 • UNKNOWN ADDRESS • Matched street and city and state • Primary number invalid • Corrected ZIP Code • Fixed abbreviations
Example 3: 1019 Northeast 70st Avenue Seattle, WA 98136	1019 NE 70th St Seattle WA 98115-5624 • Found 1 Valid Address • Matched street and city and state • Confirmed entire address • Corrected ZIP Code • Fixed city and state spelling • Address Component was changed

31 Examples were produced from the SmartyStreets website by manually entering the incorrect address, submitting it, then viewing the corrected output. https://smartystreets.com/products/single-address.

Address Correction

Address cleansing does not ensure the IIS has the most up-to-date addresses for individuals. For that, the IIS needs to use an address correction service. The USPS maintains the National Change of Address (NCOA) database, which contains the new addresses of businesses, individuals, and families who have filed a Change of Address (COA) form with the USPS within the last four years. Going to the USPS directly or through a third-party vendor, a mailing entity, such as an IIS, can obtain automated address correction services. These services provide the mailer with data files containing change-of-address and undeliverable-as-addressed information. When an IIS contracts with the USPS or a vendor, its mailed reminder-recall notices (and other mailings) will generate person- or household-specific corrected addresses. Updating addresses in this way can ensure that mailings go to the correct households, as well as help with IIS deduplication efforts and vaccination coverage analysis.

Many IIS now use an address cleansing or address correction service.[33] A description of a service sponsored by AIRA is included in Appendix F, along with three examples of IIS that use reminder-recall programs to improve address data quality.

Address Correction Service (ACS) is a mailing service that allows mailers to receive change-of-address and other reasons for non-delivery electronically and reduce the number of manual (hardcopy) address notifications. ACS has many benefits over manual address corrections. It is a cost-effective, efficient means of obtaining accurate COA or reasons for non-delivery information. Utilizing ACS can:

- Reduce manual address costs.
- Reduce labor-intensive address change functions by eliminating returned mail.
- Choose when to receive fulfillment of time-sensitive information on a daily, weekly, or monthly schedule.
- Retrieve address change information electronically via a secure Internet site to allow automated updating of mailing lists.
- Receive notifications that relate specifically to their address files.[32]

32 USPS Address Correction Service webpage. https://ribbs.usps.gov/index.cfm?page=ACS.
33 AIRA. Immunization Information System (IIS) Implementation Guidance for a Shared Address Cleansing and Geocoding Service. May 2017. http://repository.immregistries.org/resource/iis-implementation-guidance-for-a-shared-address-cleansing-and-geocoding-service/.

Patient Status Completeness and Accuracy: Improving Documentation at the Geographic Level

Appropriate Patient Active/Inactive Status (PAIS) designation in the IIS is essential to the overall quality of IIS data. The *2015 MIROW Patient Active/Inactive Status (PAIS)* best practice guide defines PAIS as "a ranking term used to describe responsibility for immunization of the individual/patient at a provider organization or geographic jurisdiction level. In other words, PAIS is a designation of the relationship of an individual/patient with a provider organization or the jurisdiction in which the individual/patient resides. PAIS at a geographic jurisdiction level conveys information with respect to the relationship of an individual to a jurisdiction."[34]

Inactive status at the geographic level should result when:

- A patient is deceased (preferably confirmed by the state Vital Statistics office).
- The current patient address is outside the jurisdiction covered by the IIS (but may still be active at the provider level if the current provider is within the jurisdiction).

Maintaining correct PAIS in the IIS improves data quality for several functions. Correct PAIS classification:

- Leads to more accurate coverage assessments (numerators and denominators derived from IIS will be more accurate)
- Improves the success of reminder-recall activities (more likely to send to patients currently residing in the jurisdiction)
- When used consistently, allows for comparability between providers within a geographic jurisdiction (because truly inactive patients will be excluded from comparison)
- Allows a geographic jurisdiction to prevent members of the community from "falling through the cracks" (even if not associated with a provider, public health agencies can follow up with patients that are active and live in their area)[35]

Patient status should be interpreted and assigned at both the provider and geographic level as data is brought into the IIS. It should also be assigned to data records at rest that lack a status or that need updating.[36] The accuracy of a status designation degrades with time. Thus, IIS staff should make special efforts to ensure that patient status remains current.

To determine if a patient has moved out of the area, the strategies for address cleansing and correction described above can be implemented. To determine if a patient is deceased, protocols should be in place that allow the IIS to import death data efficiently and mark patient status as permanently inactive. This is especially important for early childhood deaths to avoid the mailing of notices such as reminder-recall to bereaved families. IIS usually work with their state or local Vital Statistics office to receive timely death records, ideally in an automated fashion.

Another strategy to consider is the permanent inactivation of individuals who have reached an improbable age. While a formal definition of "improbable age" is not currently available, the oldest living person in the United States at the time of this writing is 114 years old, and there are only a handful over the age of 110 nationally. Thus, an IIS could reasonably inactivate records of individuals whose age exceeds 115 years. IIS staff could also compare the age distribution in the IIS to census estimates to determine if an unreasonably high proportion of individuals fall into an older age group and then take appropriate action. For further details on PAIS, the reader should consult Appendix D.

34 AIRA. Management of Patient Active/Inactive Status in Immunization Information Systems: Replacement of 2005 Guidelines. Full guide, p 67. http://repository.immregistries.org/resource/management-of-patient-active-inactive-status-in-immunization-information-systems-1/.
35 AIRA. Management of Patient Active/Inactive Status in Immunization Information Systems: Replacement of 2005 Guidelines. Mini-guide, p 31. http://repository.immregistries.org/resource/management-of-patient-active-inactive-status-in-immunization-information-systems-1/.
36 AIRA. Management of Patient Active/Inactive Status in Immunization Information Systems: Replacement of 2005 Guidelines. Full guide. http://repository.immregistries.org/resource/management-of-patient-active-inactive-status-in-immunization-information-systems-1/.

Patient and Vaccination Record Uniqueness: Improving Deduplication Processes

Duplicate record resolution is related to the data quality dimension of uniqueness. According to the DAMA UK Working Group on Data Quality Dimensions, uniqueness means that no event, person, or data element is recorded more than once based upon how that entity is identified.[37] Ideally, in an IIS, there is one demographic record for each person and one record for each vaccination event. While most IIS have processes in place that try to match incoming records with existing records, an ongoing retrospective review for duplicates is still an essential data quality activity. Back-end processes that review data at rest for duplicate records are important. Although a how-to document on deduplication is out of scope for the guide, it is crucial to acknowledge the impact of duplicate records on all dimensions of data quality:

- **Accuracy** – The IIS cannot correctly evaluate a vaccination history if two or more patient records have been incorrectly merged or kept incorrectly separate. The clinical impact is a forecast that is likely inaccurate.
- **Timeliness** – If potential duplicate patient or immunization records are not resolved quickly, the IIS will contain fragmented, incomplete records. Delays in resolution affect the IIS's ability to present an up-to-date vaccination record to providers.
- **Completeness** – Although possibly appearing complete, a vaccination event in an incorrect merge actually may contain conflicting and inaccurate data elements, and the vaccination record may contain vaccinations not actually given to the patient at hand.

> **Uniqueness:** No event, person, or data element is recorded more than once based upon how that entity is identified.

Clearly, the deduplication and combining of records must be done with the utmost care to avoid compromising validity and accuracy. Good tools for deduplication are essential at both the patient and vaccination-level. Resources and a summary of common techniques for patient-level and vaccination-level deduplication can be found in Appendix C.

Select Situations Related to Birth Data Record Management

IIS rely on birth data from Vital Records to help populate their systems. Birth data is an essential source of information used to populate the IIS, providing demographic data and often data on birth doses of hepatitis B and HBIG vaccinations. Nonetheless, a number of data quality issues can arise specific to this source. Issues arise related to adoption, amended birth certificate records, multiple births (such as twins), and birthing facility submissions to the IIS, among others. While each topic merits discussion, this guide covers two specific topics: (1) duplicate records that result from birthing facility naming conventions and (2) bad merges that result from multiple births.

Duplicate records resulting from birthing facility naming conventions

In some areas, birthing facilities submit birth records to their IIS. This data often enters the IIS via HL7 real-time interfaces and precedes birth data from Vital Statistics. This can cause data quality issues because newborns might not yet have a legal name available for submission by the birthing facility. Instead, the hospital may

37 DAMA UK Working Group on Data Quality Dimensions.

have a practice of using substitute names, such as Baby Boy or Baby Girl, on the submitted record. These names may be different enough that a match does not occur when the actual Vital Statistics birth record is received.

By the time a child reaches a few weeks of age, most IIS can reasonably expect to see data from Vital Statistics and one other provider on a newborn's record. Infant records missing a submission from Vital Statistics may be "orphaned," i.e., permanently separated from a more complete and accurate record for the same person, especially when the first-name field is suspect. Searching for records that contain only birthing facility submissions, with no vital records and no other provider data, can identify these "orphaned" records. Also, searching the IIS for obvious fake names such as Baby Girl or Baby Boy can help in identifying possible duplicates. However, new substitute fake names may arise at any time, and keeping up with them can be very time-consuming. Once such records are identified, it is appropriate to inactivate the false record or merge it with the correct record. Preventing the problem is highly desirable. Some IIS have chosen not to accept newborn data from hospitals, and this may be the optimal solution. However, if a hospital uses state-supplied vaccine for the birth dose of hepatitis B or HBIG and the IIS for vaccine tracking, problems with vaccine accountability processes can occur. It may be necessary for the IIS staff, together with the Vaccines for Children (VFC) program, to create another strategy for the hospital to track vaccine inventory.

Bad merges resulting from multiple births and other special situations

Bad merges of records can be frustrating, especially for staff who manually separate the merge only to find it recurs the next time the record is submitted. This often happens with multiple births

due to their matching demographic data for everything except the first name. The problem is compounded when the patient records contain similar first names for the children. Using a multiple birth order field identifying a child as 1 of 2 or 2 of 2 can help, except that different providers may submit different numbers for each child, compounding the problem. Even siblings of different ages may be inappropriately merged, especially if they have similar names or birth dates that partially match. Providers are often the first to notice the bad merge when viewing IIS records. IIS staff can help avoid the problem by routinely running queries that search for records with more than one birth record number. Even better, an automated program can be created to search for these problems. An improbably high number of vaccination events on a patient record is another identifier of possible bad merges, as described in Table 6. Ideally, match algorithms can be designed that consider these situations to avoid the time-intensive process of manual record separation. In addition, the automated process should include a way to flag records that have been unmerged once so they will not be merged again.

SECTION IV. IMPLEMENTATION CONSIDERATIONS AND PROCESSES

This section provides both general and specific recommendations on developing and implementing a data at rest quality plan. The recommendations are based on data quality principles and theory as well as the real-life experiences of IIS staff.

General Recommendations

As IIS staff consider implementing a data at rest quality improvement plan, many factors should be considered. These include the following:

- Start with ensuring that onboarding and incoming data analyses are done right. If resources are limited, prioritize the onboarding data validation process to prevent bad data from getting into the system (as described in the Phase One guide). Also, prioritize routine reviews of incoming data to correct issues at the provider level early on (as described in the Phase Two guide). Strict data validity standards alongside HL7 rules are important. Accepting data without such standards transfers the problem from the data source to the data consumer.
- If an IIS lacks a clear idea of the types of data quality problems that exist:
  - Analyze a large and broad amount of data, selecting high-priority metrics from Section III of the guide to gain an overall idea of quality issues.
  - Use the Immunization Information Systems Annual Report (IISAR) results to identify issues specific to your IIS.
  - Consult immunization program staff who interact with IIS data, such as epidemiologists, VFC enrollment and quality assurance staff, and those overseeing vaccine ordering, to gather their IIS data quality challenges.
- Prioritize review of recent data, as it may be easier to analyze and fix than older data. Then, analyze older data as time and resources allow. Compare results from older data to that of newer to gain insight on the evolution of problems.
- Take advantage of the Pareto principle to troubleshoot data quality issues.[38] For data quality, this principle means that quality issues are rarely uniformly distributed and that the majority of issues are caused by a minority of factors or agents. When there are many problems, statistical charts like the Pareto bar chart can help to spotlight those that are most significant. They can track data problems, speed the identification of root causes, and aid in resolution.[39]
- Create a list of triggers that will prompt the IIS to run specific data quality reports. Triggers can include:
  - IIS software upgrade or change
  - EHR software upgrades
  - Change in vaccine product availability—products no longer on the market should not show up
  - Change in vaccine codes
- Create a data quality plan for data at rest.
  - Use the step-by-step guide below to assist in the process.

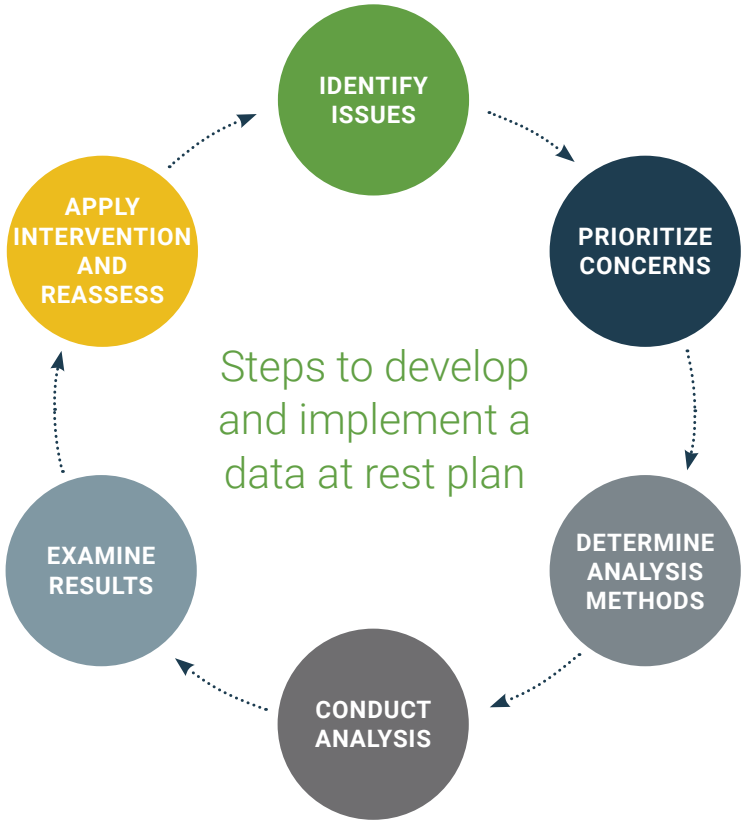
38 A process improvement technique: "The principle, named after 19th century economist Vilfredo Pareto, suggests most effects come from relatively few causes; that is, 80% of the effects come from 20% of the possible causes. One of the 'seven tools of quality.'" See https://asq.org/quality-resources/quality-glossary/p.

- Determine frequency of reports/analyses based on:
  - importance and severity of issues (determined by impact on IIS, providers, and immunization program)
  - frequency of occurrence of specific issues
  - resources and staff capacity available
- Expect the priority items to evolve and change over time.

- Track progress and change over time.
  - Use the IISAR to track data points from year to year to observe trends.
  - Monitor, store, and track results of specific DQ reports, selected by the program's priorities, so that improvements and changes can be noted.

Developing and Implementing a Data at Rest Quality Analysis Plan

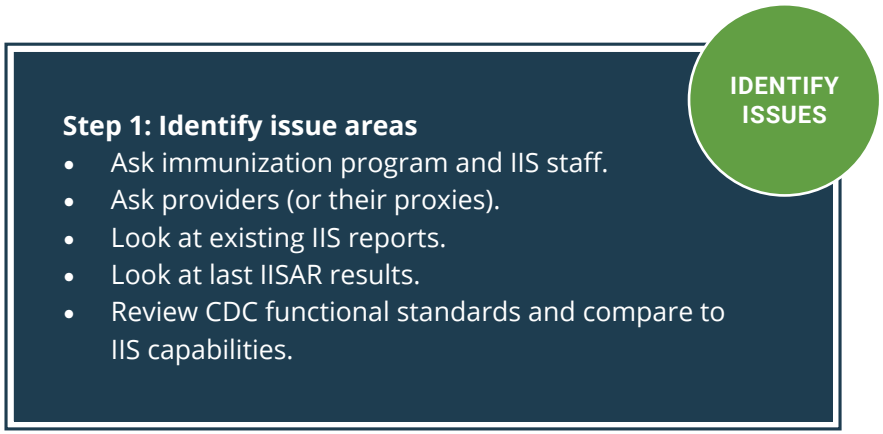
In this section, we look at the specifics for developing and implementing a quality analysis plan for data at rest. We discuss ways to identify and prioritize potential areas of concern. We provide suggestions for conducting the analysis as well as the development of a response plan and intervention. It is important to note that this process continues in a cycle and is not necessarily linear.



39 See http://asq.org/learn-about-quality/cause-analysis-tools/overview/pareto.html for more information on Pareto Charts.

Step 1: Identify data quality issue areas

Data quality means different things to different users of the IIS data. It is therefore important to get input from a variety of data consumers as the very first step to developing a data at rest quality plan. IIS users with a stake in data quality include immunization program staff, IIS staff, local health department staff, and provider-based users of the IIS. A written survey to identify problem areas may be helpful, as may be meetings with staff for more detailed discussion and brainstorming. A review of the CDC IIS Functional Standards document can also help identify areas of need. Data quality concepts are woven throughout and are reflected in multiple goals in this document.[40]



The selection of appropriate staff to interview or survey depends on the programmatic priorities and functions of each immunization program. VFC and AFIX staff usually have a stake in specific areas of IIS data quality. For VFC staff, the interest may include the accuracy of lot numbers and doses administered, child eligibility for VFC, and other fields necessary to meet VFC, state, and local requirements for vaccine tracking and inventory management. Staff conducting AFIX visits may prioritize the accuracy and completeness of vaccination data for young children in order to get the most accurate coverage results for providers. Teams that conduct reminder-recall mailings will have an interest in up-to-date addresses as well as complete vaccination histories. An IIS may have staff dedicated to data quality or whose duties include one or more aspects of data quality. All IIS staff with knowledge of data quality issues should be consulted. Examples include onboarding staff, individuals who do manual deduplication, IIS help desk personnel, and certain IT staff.

To define and scope the data quality issue areas appropriately, basic descriptive statistics may be generated at this stage. For example, does the issue occur equally across provider types? Does the average HL7 message count vary by electronic medical record type? Does the average shot type vary by type of vaccine ordered from public inventory? Does the issue occur with all shot type Code for Vaccine Administered (CVX) codes or only with some CVX codes for that shot type?

Existing reports generated by the IIS can facilitate the process of identifying the problem. The IISAR, for example, is a report that every IIS must complete annually for CDC. IIS staff can look back at results and view performance over time in various areas. Table 1 and Table 2 make note of indicators included in the IISAR that are related to data element completeness. IIS staff should exercise caution drawing inferences from IISAR measures year to year, as the questions, their placement, and restriction criteria change from time to time. An IIS may also have canned or automated reports readily available that relate to data quality. If these existing reports cannot produce IIS-wide results, they may be available at the provider level, and systemic problems can be extrapolated. See Appendix G for examples from Envision Technology, CDC, and Michigan (MCIR).

40 Immunization Information System Functional Standards, Version 4.

Step 2: Prioritize and scope quality concern areas

Once data quality concerns have been identified, they need to be prioritized. Data items critical to business operations and required reporting will typically be ranked very high. These include issues with an impact on clinical practice or federal/local requirements. For example, clinical practice and patient care are affected if the forecast is not correct. An incorrect forecast results if the vaccination history, in combination with the birth date, is incorrect or incomplete. This can result in the under-immunization of an individual. Thus, data issues that affect the forecast may take top priority. Data quality standards/thresholds for clinical decision support may be stricter than the standards for surveillance activities, since clinical decisions are made at the individual level and surveillance activities are performed at the population level.

PRIORITIZE CONCERNS

Step 2: Prioritize and scope concern areas

- Clinical practice (such as forecast accuracy).
- Public health/surveillance capability.
- Ability to meet state or federal requirements (such as VFC).
- Areas not routinely examined in the incoming data review.
- Areas you can control.

Federal or local reporting requirements related to vaccine distribution may also be a high-priority area. Federal requirements for reporting on the VFC program include reports on vaccine distribution, tracking, and accountability. Program staff also will be very interested in the accuracy of vaccine administration data, including lot number, NDC, and source of funding at the dose level. The completeness and timeliness of this data is important so they can meet deadlines for reporting and ordering as well as ensure that vaccine inventory is accurate. Another factor in prioritization is whether specific data issues are routinely examined in the incoming data review process. IIS staff should review the Phase Two guide indicators to identify missing processes for incoming data, as well as issues that are not obvious until the data is "at rest." For example, a record is subject to matching with another record at any point in time after it is in the system. When a new record coming in is matched to an existing record, the new record's data may be added to the existing record. The newly merged record may then present vaccination scenarios that were not apparent before, such as minimum interval violations that may be related to either data accuracy or clinical validity.

To refine parameters further, the team may want to consider the following questions:

- Is there a particular age range to focus on, or is it necessary to include the entire IIS population?
- Is there a need to compare results for different age groups?
- Will the analysis put limits on the geographic area covered or be broken down by area?
- Is there a need to break out data for certain vulnerable populations?
- Is the record submission date relevant? Should a range of submission dates be a parameter?
- Is the analysis limited to certain vaccine types?
- Is the EHR type of interest? Does the problem vary by EHR type?
- Is there a need to delimit or sort results by provider or practice type?

The result of Step 2 is a list of prioritized data quality issues, with problem definitions and enough descriptive details to proceed to developing the data analysis process. The specific purpose of the improvement activity should be clearly stated for each issue.

Step 3: Plan the analysis

With a particular purpose in mind, an appropriate team can be convened to give input on the analysis and approach to be used. The team should include relevant IIS and immunization program staff as well as technical and/or vendor staff. A review of the data quality indicators described in the guide can help in determining appropriate metrics and processes to use. Basic descriptive statistics generated during problem definition may help IIS organize the analysis and choose appropriate analysis tools and methods.

DETERMINE ANALYSIS METHODS

Step 3: Plan data quality analysis

- Review purpose.
- Review and confirm or revise metrics and parameters to use, e.g., population segment.
- Brainstorm approaches for analysis.
- Determine feasibility of each approach.

For each data quality dimension and its indicators, values or ranges should be established representing "good" and "bad" data quality.[41]

In determining the method of analysis, the team should review existing reports in the IIS to see if something already exists that could be used or modified. Options might include:

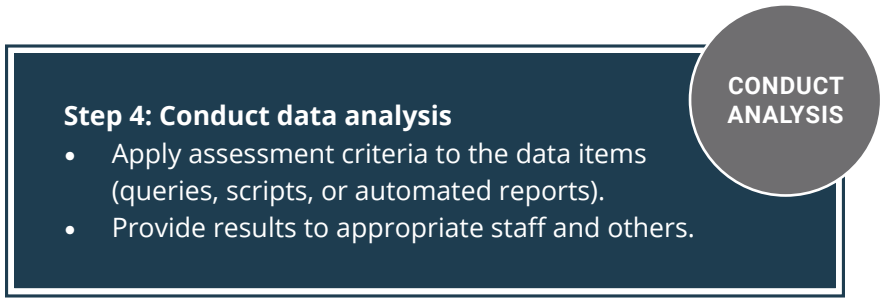
- Use of existing report or query
- Modification of existing report or query
- Development of new report or query

The feasibility of any given method will depend at least in part on the resources needed: time, staff, and financial. If the required resources are prohibitive, it may be necessary to rethink the approach and look for other methods, or interim steps, that could be taken at a lower cost, such as doing the analysis in stages. Finally, it may be clear that a given measure is not feasible at the present time and needs to be postponed.

41 DAMA UK Working Group on Data Quality Dimensions.

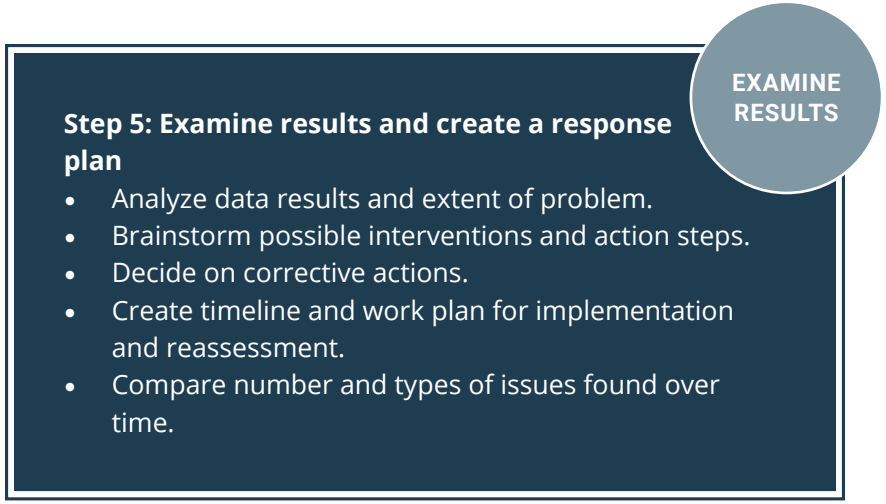
Step 4: Conduct data analysis

The next step is to conduct the analysis by applying the assessment criteria to the data items through a script, query, or other report format. However, before conducting the data analysis, it is advisable to determine the demands it will place on the system and to find out from information technology staff if alternative systems are available for the analysis to minimize impact. Some database systems are designed with a data warehouse or operational data store that allows access to an exact copy of the raw IIS data. These systems are a facsimile of the IIS, so large analyses should not impact the actual IIS. But some knowledge of SQL may be required depending on the interface client. If such systems are not in place and large data sets will be reviewed, it is important to find ways to mitigate the impact on regular operations. The process may need to be conducted in off-hours, or other special arrangements may need to be made. Having the right staff involved will help ensure there are no unintended consequences.



Step 5: Examine results and create a response plan

With the data analysis output in hand, the team can review the results and determine if data quality is acceptable or not. If not, the significance and extent of the problem can be further examined. The team can explore corrective actions to clean the data and improve processes to prevent future recurrences. A corrective action plan should be created, with clear objectives and timelines.



As the results are reviewed, the team may ask the following questions:

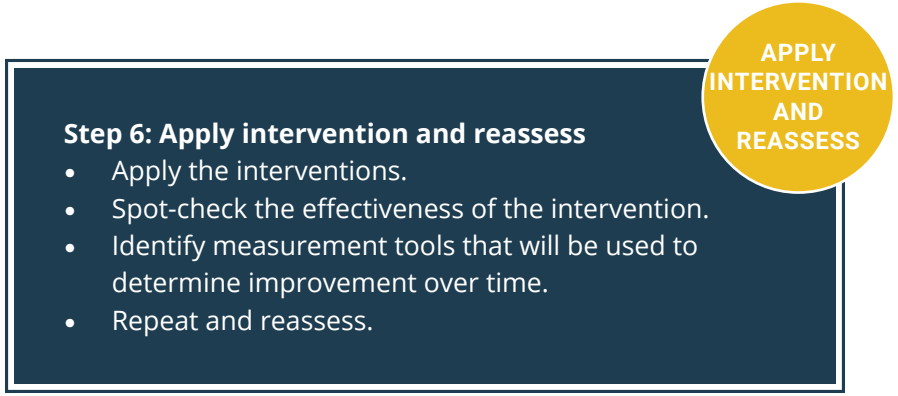
- How long has it been since the data was reported?
  - If the data was reported a long time ago, data quality issues may be harder to fix since the submitting system's logs that reported the data might have been overwritten. The practice may have also changed their EHR vendor.
- Did the error occur before or after the IIS controlled the data?
  - If the error occurred before the IIS was in control of the data, the provider should be encouraged to resend the corrected data.

  - If the error occurred after the IIS was in control of the data, is it something that can be prevented in the future?
- Does something in the data need to be fixed or corrected? If so, can it be fixed?
  - What are the options for fixing it? Is there an automated script that can be applied? Do providers need to make the changes? Will it require IIS staff time to manually correct the data?
  - Are resources available to make the corrections?
  - How will corrections be documented?
- If data cannot be fixed or corrected, does it need to be masked or given conditional access to prevent incorrect clinical decisions or surveillance activities?
- Were the parameters adequate to catch all the problems?
  - Is additional data needed? For example, should the parameters be expanded or changed to look for more errors?
  - Would it be helpful to drill down into the data to get at provider-level results or to change the age group or expand the data submission date range?
- Do changes need to be made to incoming data processes to prevent the problem from recurring?
  - Is provider-staff training needed to prevent recurrence of the issue?
  - Would a programming change prevent the problem in the future (e.g., more restrictive data submission standards)?

This is a good time to determine the level of information that needs retention. The decision to retain the results information should be based on the need or desire to evaluate data quality changes over time. Ongoing measurement tools and the frequency of their application should be documented.

Step 6: Apply intervention and reassess

After applying the intervention, it is advisable to spot-check the effectiveness of the intervention—perhaps a manual look at 10-15% of the affected records to see if the appropriate changes resulted. If the intervention appears to have taken care of the problem, a follow-up assessment can be scheduled to occur at a certain point in the future to make sure the problem has not recurred.



SECTION V. CONCLUSION

AIRA's three-guide series on IIS data quality practices offers practical strategies for improving IIS data. The first guide focused on the validation of data during the onboarding phase of a new electronic data exchange (EDE) source. The second guide covered ongoing monitoring and review of incoming data at the provider level. This third guide completes the series by identifying and highlighting data quality practices for data at rest.

All three phases of data quality examination are important for ensuring high-quality data. The intent of the guides is to give IIS useful resources to build or expand their data quality practices. The intent is not to describe a one-size-fits-all approach. Recommendations made in the guide are independent of particular IIS implementations and technology solutions. They can support the wide variety of IIS implementation and technology strategies.

While the first two guides in the series focused on provider-specific data quality, the data at rest guide allows for the examination of data across providers, looking for issues that may be specific to other parameters, such as age groups, vaccine types, or dates of submission. Consistent with the other guides, the data at rest guide explores the dimensions of accuracy, completeness, and timeliness. In addition, the guide examines the dimensions of consistency, validity, and uniqueness.

The data at rest guide allows for the examination of data across providers, looking for issues that may be specific to other parameters, such as age groups, vaccine types, or dates of submission.

These dimensions are of special interest to IIS because of their implications for:

- Clinical decision making
- Vaccine tracking and accountability
- Vaccination coverage assessments
- Public health research

AIRA developed the guide with the expectation that each IIS program will adjust the implementation of measures and practices to its own specific needs. Decisions on what to assess should be based on multiple factors, including IIS-specific data quality concerns, current and planned data use, and program capacity. That said, indicators that affect clinical care (e.g., accuracy of vaccine type and date of administration) merit a high priority. Likewise, indicators related to federal or local requirements, such as VFC accountability rules, are often of high priority to an IIS. The list of recommendations presented in the guide is not exhaustive. Individual IIS may choose to implement additional rules based on their unique requirements. Accordingly, Section IV provides a template—a step-by-step process—that IIS staff can use to create a data at rest quality plan that meets their specific needs.

APPENDICES

APPENDIX A – ABBREVIATIONS

ABBREVIATIONS	
ACIP	Advisory Committee on Immunization Practices
ACS	Address Correction Service (a term of the USPS)
AFIX	Assessment, Feedback, Incentives, and Exchange
AIRA	American Immunization Registry Association
ASC	Assessment Steering Committee
BR	Business Rule
CASS	Coding Accuracy Support System (a term of the USPS)
CDC	Centers for Disease Control and Prevention
CDS	Clinical Decision Support
CDSi	CDC Clinical Decision Support Logic for immunizations
COA	Change of Address
CVX	Code for Vaccine Administered
DOB	Date of Birth
DQA	Data Quality Assurance
DQ Guide	Data Validation Guide

ABBREVIATIONS, CONT.	
EDE	Electronic Data Exchange
EHR	Electronic Health Record
HBIG	Hepatitis B Immune Globulin
HL7	Health Level Seven
IIS	Immunization Information System
IISAR	Immunization Information Systems Annual Report (a CDC report required of each IIS annually)
MIROW	Modeling of Immunization Registry Operations Workgroup
NCOA	National Change of Address (a term of the USPS)
NDC	National Drug Code
NPI	National Provider Identifier
SME	Subject Matter Expert
USPS	United States Postal Service
VFC	Vaccines for Children
VIS	Vaccine Information Statement

APPENDIX B – GLOSSARY OF TERMS

ACCURACY – A dimension of data quality. Refers to the degree to which the data reflect reality. In the case of immunization data, accuracy refers to the degree to which the vaccination data match the clinical encounter and the demographic data match the person.

CLINICAL DECISION SUPPORT (CDS) [FOR IMMUNIZATIONS] – An automated process that determines the recommended immunizations needed for a patient and delivers these recommendations to the health care provider.[42]

CODE FOR VACCINE ADMINISTERED (CVX CODE) – A numerical code that describes a vaccine type. CVX codes are assigned by CDC to support electronic messaging of immunization histories via HL7.

COMPLETENESS – A dimension of data quality. Refers to the degree to which full information about a data set or an individual data element is captured in the IIS. In the case of data submissions to an IIS, completeness refers to the submission of all relevant data from the submitters and to the completeness of individual data elements of interest.

CONSISTENCY – A dimension of data quality. Refers to the absence of difference when comparing two or more representations of a thing against a definition or against itself.

ELECTRONIC HEALTH RECORD (EHR) – System utilized by the provider organization. EHR generally refers to the technology and all the software of an electronic recordkeeping system used in health care. Electronic medical record refers to the medical records maintained in an EHR system.

GEOCODING – The process of transforming a description of a location—such as a pair of coordinates, an address, or a name of a place—to a location on the earth's surface. You can geocode by entering one location description at a time or by providing many of them at once in a table. The resulting locations are output as geographic features with attributes that can be used for mapping or spatial analysis.

HEALTH LEVEL SEVEN (HL7) – A nationally recognized standard for electronic data exchange between systems housing health care data.

INTERFACE – The electronic connection between the IIS and sources of immunization data, such as EHRs, Vital Statistics, and others, for electronic data exchange between these systems.

LOT NUMBER – The number assigned by the manufacturer to a specific batch of vaccine product type. Lot number can be used by IIS to track administered vaccines.

LOT NUMBER EXPIRATION DATE – The expiration date assigned to each lot of vaccine by the manufacturer. Beyond this date, the vaccine should no longer be administered.

MANUFACTURER (MVX CODE) – Manufacturer refers to the organization that manufactures a specific vaccine. MVX codes are assigned by CDC to support electronic messaging of manufacturers via HL7.

42 The CDC CDSi Logic Specification and Supporting Data are available at https://www.cdc.gov/vaccines/programs/iis/cdsi.html.

MINIMUM/MANDATORY DATA SET – Minimum requirements for data elements to be submitted to an IIS, without which a data submission should be rejected, per AIRA document Chapter 5: Updates for 2008 MIROW DQA Guide, Business Rules 104 and 105.

BR 104 – DEMOGRAPHIC ONLY
IIS-AO ID (Recorder)
Patient Date of Birth
Patient Name, First
Patient Name, Last
Birth Certificate Number (if from Vital Statistics)
Birth Facility (code, name, address) (if from Vital Statistics)
Gender

BR 105R1 ADMINISTERED VACCINATION EVENT	BR 105R2 "HISTORICAL" VACCINATION EVENT
IIS-AO ID (Vaccinator/Recorder)	N/A
Patient Name, First	Patient Name, First
Patient Name, Last	Patient Name, Last
Patient Date of Birth	Patient Date of Birth
Vaccination Encounter Date	Vaccination Encounter Date
Vaccine Type	Vaccine Type
Administered/Historical Indicator = "Administered"	Administered/Historical Indicator = "Historical"
Lot Number	N/A

NATIONAL DRUG CODE (NDC) – NDC is defined as a unique numeric identifier of the vaccine product type. Each drug product is assigned a unique three-segment number. This number, known as the NDC, identifies the labeler (manufacturer or distributor), product, and package.

NATIONAL PROVIDER IDENTIFIER (NPI) – NPI is a unique numeric identifier issued by the Centers for Medicare and Medicaid Services used to identify health care providers.

ONBOARDING – Process of bringing a new data exchange source from first contact to going live with the exchange. Can also apply to the process of enhancing or changing an existing data interface.

PATIENT RECORD DEDUPLICATION / PATIENT-LEVEL DEDUPLICATION – The process of identifying redundant patient records in a database and consolidating or linking duplicate patient records of the same individual.

PROVIDER ORGANIZATION – An organization that provides vaccination services or is accountable for an entity that provides vaccination services. A provider organization can be a solo practice with one clinical site or can contain a collection of related providers (e.g., clinicians, physicians, nurses) with multiple sites.

TIMELINESS – A dimension of data quality; refers to the length of time between an event of interest (e.g., vaccination) and the event's capture in the IIS. A timely record is one that was captured in the IIS within recommended limits.

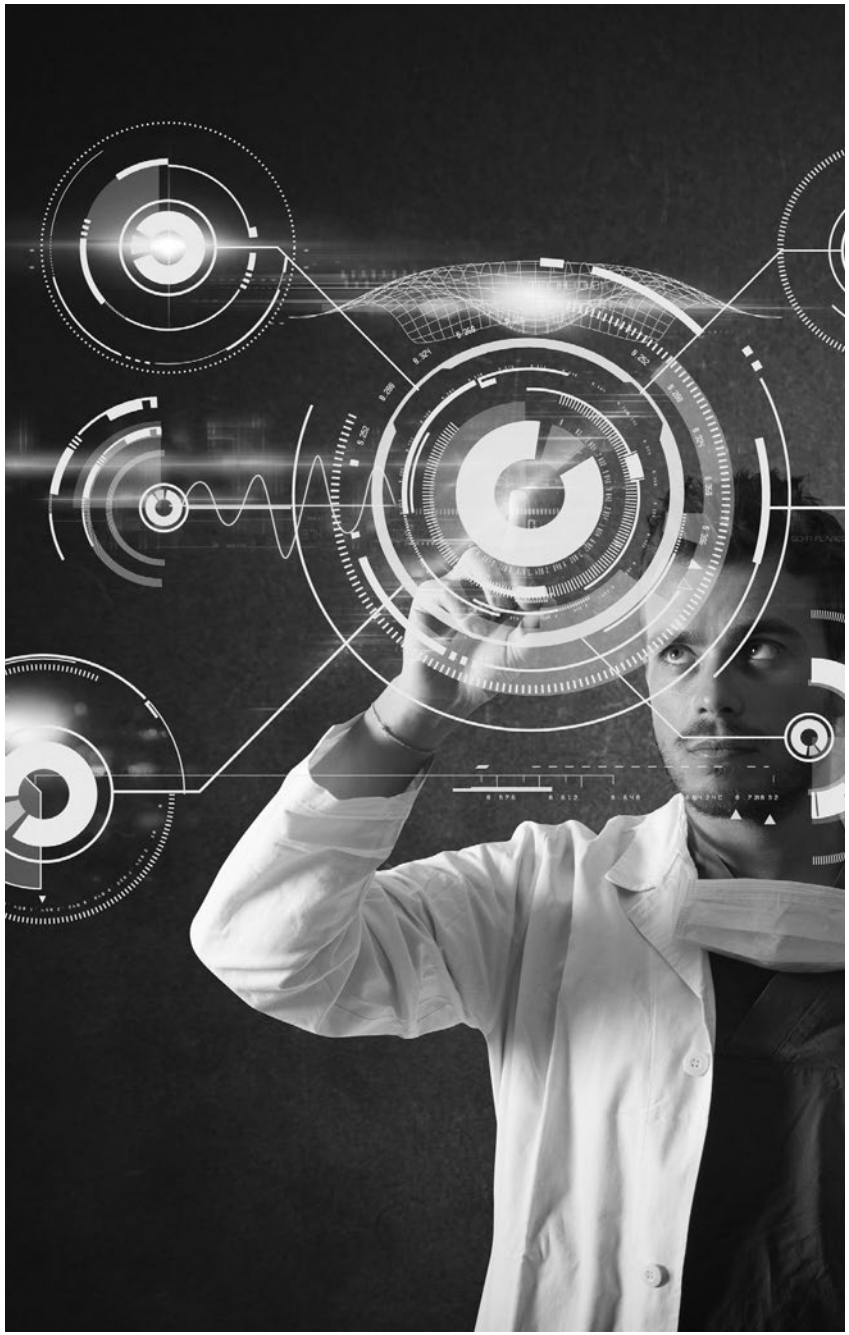
TRADE NAME – Indicates the manufacturer's proprietary name for a product, and in some cases, its intended use (e.g., adults, pediatrics) is included in the name.

VACCINATION ENCOUNTER DATE – Same as vaccination administration date.

VACCINE EXPIRATION DATE – This is the expiration date assigned to each lot of vaccine by the manufacturer. Beyond this date, the vaccine should no longer be administered.

VALIDITY – A dimension of data quality; refers to the degree to which the data conforms to rules of what is accepted or expected by the IIS. May be measured differently if considering clinical validity (e.g., minimum interval violations) as compared to technical validity (e.g., conformance to HL7 syntax).

UNIQUENESS – A dimension of data quality and a component of accuracy. Refers to a singular recording of an event, person, or data element, nothing being recorded more than once based upon how that entity is identified.



APPENDIX C – ADDITIONAL RESOURCES FOR PATIENT-LEVEL AND VACCINATION-LEVEL DEDUPLICATION

Patient-Level Deduplication Resources

A few resources for more information on patient-level deduplication are described below:

CDC. Immunization Information Systems Patient-Level De-Duplication Best Practices.[43]

Excerpt from Section 4.2. Retrospective Review:
"IIS retrospective processing examines the existing records in an IIS database checking for duplicates. Retrospective patient deduplication (also referred to as "back-end" deduplication) involves the practice of looking at the data in an IIS database and then determining if duplicate patient records exist. The objective of back-end or retrospective review is to identify and resolve duplicate patient records that represent errors in the IIS database. It can also be used to check IIS data quality.

Numerous techniques can be used for retrospective review. These techniques include database queries, manual flagging, and "spiders" walking through a database to check for and flag potential duplicate records. Most commonly, a prescheduled or on-demand batch retrospective patient deduplication process is utilized. Automated retrospective deduplication processing can be very valuable in preserving overall data quality.

The advantage of a retrospective back-end approach is that the system can automatically examine the entire database and potentially find and resolve large numbers of exact duplicates without any human intervention. Accordingly, retrospective processes can be very valuable for integrated systems where large numbers of records are coming from multiple sources.

The disadvantage of a back-end approach is that it can result in the need for extensive manual record reviews. Researching the records to make a matching determination can be time-consuming and costly.

Once records have been adjudicated through manual review, functionality should exist to retain a record of the adjudication. As a best practice, retrospective deduplication processing will contain the information needed to avoid having to reconsider previously adjudicated records.

Thresholds and decisions regarding incomplete information as well as other types of decisions need to be formally documented.

43 Immunization Information Systems Patient-Level De-Duplication Best Practices. National Center for Immunization and Respiratory Disease (NCIRD) Immunization Information Systems Support Branch (IISSB), June 25, 2013. https://www.cdc.gov/vaccines/programs/iis/interop-proj/downloads/de-duplication.pdf pp. 33-34.

Retrospective Processing Excerpt[44]

A record is selected from the IIS database.	Each record in the database is checked against similar blocked records.
Potential matches for this record are found based upon a selection and blocking criteria.	Because these processes can run in the background, their logic can be very extensive. Additionally, different batch processes can be designed to check for specific types of suspected problems.
Pairs of records are evaluated to determine if they are duplicates. Based upon this examination, records may be declared a match and combined, declared a non-match with no other action taken, or be written to a pending file for further human manual review.	Records that have been previously adjudicated can be written to a table in such a manner that they never have to be compared again.

CDC. Immunization Information Systems Patient-Level De-Duplication Test Case Development, Testing & Utilization.[45]

Excerpts:
"Patient-level deduplication testing is an important area. The strategic employment of test cases can help systematically improve IIS operations, improve the sensitivity and specificity of patient deduplication processing, reduce the need for manual reviews, and meet the objective for greater automation and data quality. Well-constructed, best-practice-oriented test cases represent a significant aid and resource to the IIS national practice community." (p. 3)

"The work products documented in this report volume include:

- An evaluation of the existing 2002 Patient-level Deduplication Tool Kit
- An examination of the IIS data used for patient deduplication testing
- An examination of the most typical and challenging patient deduplication problems
- A review of complex test cases and deduplication problems where two or more variables may cause identity ambiguities
- Important discussions around the evolution of IIS patient deduplication testing needs
- Consensus views regarding the construction of new and useful deduplication test cases
- Observations around the testing 'roadmaps' needed to improve IIS patient deduplication capabilities, reduce the need for manual record reviews, and drive greater collaboration and standardization in the national practice community

In addition to this report, the panel developed a Test Case Matrix that contains the actual evolved test cases and tools for utilizing them." (p. 5)

44 Table C-1 is part of Section 4.2 within the quoted document.
45 Immunization Information Systems Patient-Level De-Duplication Test Case Development, Testing & Utilization. National Center for Immunization and Respiratory Disease (NCIRD) Immunization Information Systems Support Branch (IISSB) Version 1.0. July 2, 2013. https://www.cdc.gov/vaccines/programs/iis/interop-proj/downloads/de-duplication-test-cases-report.pdf.

The Public Health Informatics Institute. The Unique Records Portfolio: A guide to resolving duplicate records in health information systems.[46]

Excerpt:
"The Portfolio addresses the challenge of duplicate records in integrated information systems by:

- Offering consistent terminology and definitions for concepts important to uniquely identifying individuals and their data contained within disparate information systems.
- Explaining the possible technical approaches (architectures) that can be used to create integrated information.
- Reviewing the benefits and limitations of each approach.
- Providing tools to help users think through the decisions they need to make (within the context of their roles) in determining how to design, implement, and use an integrated system." (p 12)

AIRA. Consolidating Demographic Records and Vaccination Event Records.[47]

Excerpt:
"The consolidating records process begins when two records are identified as matched records during deduplication. The document does not cover the full details of the deduplication process; rather, the focus is only on the consolidating records piece. Matched records are two records that represent the same patient or the same vaccination event. Information from the two matched records should be combined into one consolidated record. Information about a patient is consolidated in a demographic record, and information about a vaccination event is consolidated in a vaccination event record. If there are more than two matched records, then the consolidation process runs repeatedly, comparing two records at a time. Once two records are matched, the best values for each data element will be selected to form a consolidated record…This document provides best practice recommendations on how an IIS should consolidate records. While viewing this document, it may help the reader to consider the following three principles, which are foundational to consolidation:

- The essence of consolidation is to select the best value for each data element from all available data sources. An IIS achieves this by comparing values from separate records for a single data element and selecting the better of the values. Via this process, an IIS distills and retains the best information into the consolidated record.
- The act of consolidation will create a new record or update an existing record.
- The functionality of consolidation relies on certain original information submitted to IIS by immunization providers and other data sources being accessible in the IIS. That data is vital for ongoing consolidation and for fixing incorrectly merged records." (p. 2)

46 The Unique Records Portfolio: A guide to resolving duplicate records in health information systems. Clyde, Salkowitz, and PHII. 2006. https://www.phii.org/resources/view/4380/unique-records-portfolio-guide-resolving-duplicate-records-health-information.
47 AIRA. Consolidating Demographic Records and Vaccination Event Records. Modeling of Immunization Registry Operations Workgroup (MIROW). August 21, 2017. http://repository.immregistries.org/resources/search/consolidating+demographic+records.

Vaccination-Level Deduplication Resources

One resource for more information on vaccination-level deduplication is described below.

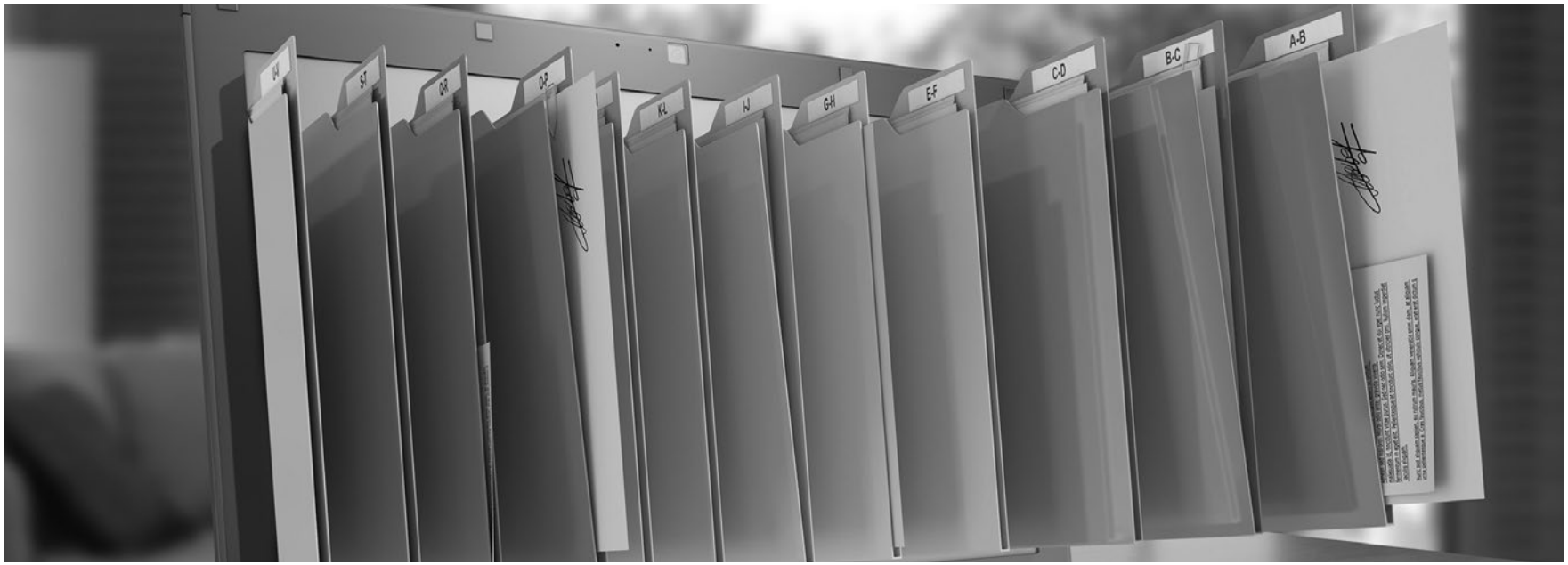
Vaccination-Level Deduplication in Immunization Information Systems.[48]

"The implementation of the best practices in this mini-guide enable IIS to develop automated algorithms that will identify potential duplicate records for vaccination events, determine which of these are indeed duplicates, identify and select the best record among duplicates and create a view of a consolidated record. The results ensure the application of a consistent, logical approach to deduplication. Doing so ultimately prevents under- and over-immunization of patients by having an up-to-date, accurate immunization record."[49]

48 AIRA. Vaccination Level Deduplication in Immunization Information Systems. Modeling of Immunization Registry Operations Work Group (MIROW). Dec 2007. http://repository.immregistries.org/resource/vaccination-level-deduplication-in-immunization-information-systems-1/.
49 AIRA. Management of Patient Active/Inactive Status in Immunization Information Systems: Replacement of 2005 Guidelines. Mini-guide, p 12. http://repository.immregistries.org/resource/management-of-patient-active-inactive-status-in-immunization-information-systems-1/.

APPENDIX D – PATIENT ACTIVE/INACTIVE STATUS DISCUSSION

According to the AIRA guide *Management of Patient Active/Inactive Status in Immunization Information Systems: Replacement of 2005 Guidelines*, inactivation of a patient record at the geographic level should occur only when definite information is available: either a death has occurred or the patient has moved out of the jurisdiction.[50] When inactive status is suspected but not actually known, the default status is "unknown." Situations falling in this category include (1) when the IIS has never received an address or vaccination information about an individual and (2) if the IIS has not received demographic and/or immunization information for an individual for an extended period of time. Although these conditions suggest that a person may no longer be in the jurisdiction, other causes could be at play (such as no vaccinations received by the patient or vaccinations received from non-submitting providers). IIS staff are encouraged to undertake analysis of patients with "unknown" status in order to get more clarity on these situations. Strategies such as reminder mailings and the address correction process described earlier in the guide could be implemented to gain clarity on patient status.



50 http://repository.immregistries.org/files/resources/5835adc2dad8d/mirow_pais_full_guide.pdf.

APPENDIX E – REFERENCES

AIRA. Analytic Guide for Assessing Vaccination Coverage Using an IIS. November 2015. http://repository.immregistries.org/resource/analytic-guide-for-assessing-vaccination-coverage-using-an-iis/

AIRA. CDC-Endorsed Data Elements AIRA Discovery Session. November 27, 2017. http://repository.immregistries.org/resource/aira-discovery-session-cdc-endorsed-data-elements/from/type:webinars

AIRA. Comparing & Communicating Vaccination Coverage Estimates from Immunization Information Systems, the National Immunization Survey, and Related Assessments. August 2017. http://repository.immregistries.org/resource/comparing-and-communicating-vaccination-coverage-estimates-from-iis-nis-and-related-assessments/

AIRA. Consolidating Demographic Records and Vaccination Event Records. American Immunization Registry Association (AIRA) Modeling of Immunization Registry Operations Work Group (MIROW). August 21, 2017. http://repository.immregistries.org/resources/search/consolidating+demographic+records

AIRA. IIS Data Quality Practices: Monitoring and Evaluating Data Submissions. American Immunization Registry Association. September 2017. http://repository.immregistries.org/resource/iis-data-quality-practices-monitoring-and-evaluating-data-submissions/

AIRA. Data Validation Guide for the Onboarding Process. American Immunization Registry Association. September 2016. http://repository.immregistries.org/resource/data-validation-guide-for-the-iis-onboarding-process/

AIRA. Data Quality Assurance in Immunization Information Systems: Incoming Data. American Immunization Registry Association (AIRA) Modeling of Immunization Registry Operations Work Group (MIROW). 2008. http://repository.immregistries.org/resource/data-quality-assurance-in-immunization-information-systems-incoming-data-1/

AIRA. Data Quality Assurance in Immunization Information Systems: Selected Aspects. American Immunization Registry Association (AIRA) Modeling of Immunization Registry Operations Work Group (MIROW). May 2013. http://repository.immregistries.org/resource/data-quality-assurance-in-immunization-information-systems-selected-aspects/

AIRA. 2017 Education Survey Summary. July 2017.

AIRA. Immunization Information System (IIS) Implementation Guidance for a Shared Address Cleansing and Geocoding Service. May 2017. http://repository.immregistries.org/resource/iis-implementation-guidance-for-a-shared-address-cleansing-and-geocoding-service/

AIRA. Lot Number Validation Best Practices. Modeling of Immunization Registry Operations Work Group (MIROW) Micro-guide. May 2014. http://repository.immregistries.org/resource/mirow-micro-guide-lot-number-validation-best-practices/

AIRA. Management of Patient Active/Inactive Status in Immunization Information Systems: Replacement of 2005 Guidelines. American Immunization Registry Association (AIRA) Modeling of Immunization Registry Operations Work Group (MIROW). April 2015. http://repository.immregistries.org/resource/management-of-patient-active-inactive-status-in-immunization-information-systems-1/

AIRA. Reminder/Recall in Immunization Information Systems. American Immunization Registry Association (AIRA) Modeling of Immunization Registry Operations Work Group (MIROW). April 2009. http://repository.immregistries.org/resource/reminder-recall-in-immunization-information-systems-1/

AIRA. Vaccination Level Deduplication in Immunization Information Systems. American Immunization Registry Association (AIRA) Modeling of Immunization Registry Operations Work Group (MIROW). December 2007. http://repository.immregistries.org/resource/vaccination-level-deduplication-in-immunization-information-systems-1/

AIRA. Vaccine Lot Number Patterns: Unit of Sale/Unit of Use Guidance. July 12, 2017. http://repository.immregistries.org/resource/guidance-on-unit-of-sale-unit-of-use-lot-numbers/

ASQ.org. http://asq.org/learn-about-quality/cause-analysis-tools/overview/pareto.html. Retrieved February 25, 2018.

Campbell, M. Personal communication with Michelle Campbell, Washington State IIS Data Quality Specialist. February 23, 2018.

CDC. CDC Endorsed Data Elements 2018-2022. https://www.cdc.gov/vaccines/programs/iis/core-data-elements.html

CDC. Epidemiology and Prevention of Vaccine-Preventable Diseases (The Pink Book). Hamborsky, J., Kroger, A., Wolfe, S., eds. 13th ed. Washington, D.C. Public Health Foundation, 2015. https://www.cdc.gov/vaccines/pubs/pinkbook/index.html

CDC. HL7 Version 2.5.1 Implementation Guide for Immunization Messaging, v 1.5, November 2014, and Addendum. July 2015. https://www.cdc.gov/vaccines/programs/iis/technical-guidance/hl7.html

CDC. IIS Functional Standards, v 4.0. January 2018. https://www.cdc.gov/vaccines/programs/iis/functional-standards/func-stds-v4-0.html

Chen, H. et al. A Review of Data Quality Assessment Methods for Public Health Information Systems. Int. J. Environ. Res. Public Health 2014, 11, 5170-5207; doi:10.3390/ijerph110505170. http://www.mdpi.com/1660-4601/11/5/5170

Clyde, S., Salkowitz, S., PHII. The Unique Records Portfolio: A guide to resolving duplicate records in health information systems. Public Health Informatics Institute. Decatur, GA. 2006. https://www.phii.org/resources/view/4380/unique-records-portfolio-guide-resolving-duplicate-records-health-information

DAMA UK Working Group on Data Quality Dimensions. The Six Primary Dimensions for Data Quality Assessment Defining Data Quality Dimensions. October 2013. https://www.whitepapers.em360tech.com/wp-content/files_mf/1407250286DAMAUKDQDimensionsWhitePaperR37.pdf

Dombkowski, K.J. et al. Assessing the burden of undeliverable immunization reminder and recall notifications. Prev Med. 2011 Dec; 53(6):424-6. doi: 10.1016/j.ypmed.2011.09.014. Epub 2011 Oct 6. https://www.ncbi.nlm.nih.gov/pubmed/22001688

Kuramoto, S., Muscoplat, M. Using Address Checking Services to Facilitate Reminder and Recall in Minnesota. AIRA National Meeting. April 15, 2015. http://repository.immregistries.org/resource/track-b-using-iis-to-improve-immunization-rates-2/

Kuramoto, S. Personal Communication with Sydney Kuramoto, MIIC Informatician, Minnesota Immunization Information Connection (MIIC). February 23, 2018.

Michigan Care Improvement Registry/MCIR Data Quality Improvement (DQI) Plan Version 2.0 June 3, 2016. Received by personal communication from Bea Salada, MDHHS state MCIR coordinator on March 1, 2018.
US Postal Service. ACS Product Information Guide. July 2, 2013. http://about.usps.com/publications/pub8.pdf

Washington State Immunization Information System and Child Profile. Health Promotion web page: https://www.doh.wa.gov/YouandYourFamily/Immunization/ChildProfileHealthPromotion. Retrieved on February 23, 2018.

Woinarowicz, M. Conducting Statewide Recall Using the NDIIS Successes, Challenges, and Lessons Learned. AIRA National Meeting. April 22, 2015. http://repository.immregistries.org/resource/track-b-using-iis-to-improve-immunization-rates/

Woinarowicz, M. Personal Communication with Mary Woinarowicz, North Dakota IIS Manager, February 4, 2018.

APPENDIX F – EXAMPLES OF IIS USING ADDRESS CORRECTION SERVICE

AIRA-Sponsored Address Cleansing Service

The SmartyStreets address cleansing shared service is available to all AIRA member IIS programs. This service is a data quality component that comes at no cost to the IIS programs and is supported through AIRA and CDC. The service provides address standardization, validation, and geocoding for all data within an IIS, including patient, responsible party, and provider/partner addresses. IIS can choose from several ways of connecting to the service, including manual batch processing, automated batch processing, an HL7 real-time solution, and/or a user application programming interface. For more information on this AIRA service, visit http://repository.immregistries.org/resource/iis-implementation-guidance-for-a-shared-address-cleansing-and-geocoding-service/.

IIS Example 1: North Dakota Immunization Information System (NDIIS)

In 2013, the NDIIS conducted a centralized, statewide adolescent recall pilot project. Based on lessons learned from a previous infant recall project, the program conducted an address cleanup prior to the mailing of recall notices. They submitted NDIIS data files containing addresses for North Dakota adolescents to the US Postal Service National Change of Address (NCOA) system and used the NCOA return data to update address information in the IIS. Those with out-of-state addresses were marked in the IIS as "moved or gone elsewhere" (MOGE). Recall letters returned as "undeliverable" were marked as "lost to follow-up." In this pilot, 4% of adolescent addresses submitted to the NCOA were newly marked as MOGE, while 0.5% were returned as undeliverable.

After the pilot, the IIS continued the adolescent recall program using a third-party vendor for the mass mailings. NDIIS provides the addresses from the recall report to the vendor. The vendor then processes the addresses through the NCOA before stamping and mailing the envelopes with the verified or updated client addresses. The vendor sends back the updated addresses to NDIIS. Initially, a staff person manually updated records in the IIS. Currently, an automated process is in place—a script is run that updates all of the addresses first and then updates records that need to have their status changed (e.g., to MOGE).

Since the initial pilot, NDIIS has expanded its recall efforts. At this time, it mails quarterly recall notices to children 24-35 months old and to teens 12-17 years old. It also sends annual reminders to kids entering kindergarten and 7th grade for school-required immunizations, as well as HPV reminders to children who recently turned 11 years of age. At least twice a year, address data is submitted directly to the NCOA for verification and updating, while the third-party vendor continues to handle all mass mailing efforts.[51,52]

51 Woinarowicz, M. Conducting Statewide Recall Using the NDIIS. Presented at AIRA National Conference, 2015.
52 Personal communication with Mary Woinarowicz, NDIIS Manager. February 4, 2018.

In summary, the project has fulfilled the important goal of reminding parents about vaccinations due for their children. In addition, it has the added benefit of improving address quality and patient status in the IIS. In the initial adolescent project, 4% of the teens were found to have moved out of state and thus were marked as Moved or Gone Elsewhere, thus providing a more accurate denominator count for the IIS population.

IIS Example 2: Minnesota Immunization Information Connection (MIIC)

The Minnesota Immunization Information Connection (MIIC) received 2013 Prevention & Public Health Funds (PPHF) from CDC to increase adolescent HPV rates. It implemented reminder-recall activities in 2014 that included a statewide postcard and regional mailings. The target group was 11- to 12-year-olds with a focus on reminders about recommended teen vaccines (Tdap, MCV, and HPV). The total number of clients in the birth-date range with Minnesota addresses was 141,183.

Prior to sending out the mailings, MIIC contracted with an address checking vendor. It utilized the vendor to update addresses and phone numbers. MIIC chose the company based on pricing and ease in contracting. The contract included a data security clause. Sources used by the vendor included utility hook-up, credit reports, and the National Change of Address service (NCOA).

A batch file exchange was used to provide data to the vendor and to receive it back into the IIS. Data elements provided by MIIC were:

- Client ID
- Date of Birth
- Name
- Current Address
- Phone
- Parent Information

Data elements for each client provided back by the vendor were:

- Match Indicator
- Current Address
- Current Phone
- Last Reported Date

The “match indicator” was comprised of a combination of mother’s and father’s name and was used to prioritize the information received back. The highest-priority data was subsequently loaded into MIIC. The definition of a match is a business decision made by the IIS and depends in part on the vendor and the system they have in place.

Of the 141,183 addresses reviewed, 30% were not found. “Not found” meant that the address checking service was unable to locate address information using the person data provided. The remaining addresses (70% of the total) were either confirmed or updated. Of these, 57% were confirmed as identical (or nearly so) to MIIC information, while 40% received major updates. Major updates were defined as situations where at least two differences among the city, zip code, and street address information existed in comparing the MIIC data to the returned address information. Minor changes, such as an incorrect street label or apartment number, comprised a very small percentage of address updates (3%).

CATEGORY	N	PERCENT
Found	99,105	70.2%
Confirmed	56,602	40.1%
Updated	42,503	30.1%
Not Found	42,708	30.1%
Total	141,183	

MIIC updated its records based on the returned information, and mailings were not sent to those with out-of-state addresses. However, mailings were sent to those "not found" by the address checking service if they still had a Minnesota address and were living when the final cohort was extracted.

MIIC continues population-based reminder-recall efforts. This is conducted mostly through its regional coordinators, with some also done at the state level. Evaluation is ongoing, including the use of a different address checking service.

Summary:

- Postcards sent to clients who were found by the vendor were approximately 75% less likely to be undeliverable than clients not found.
- Compared to clients who were not found by the vendor:
  - Postcards sent to clients who had confirmed address information were approximately 91% less likely to be undeliverable.
  - Postcards sent to clients who had updated address information from the vendor were approximately 52% less likely to be undeliverable.
- Postcard return rates show a slight positive association with poverty rate at the zip code tabulation area level.

MIIC concluded that:

- The vendor's address services provided new information for 70% of the cohort.
- Use of address checking services shows promise in reducing undeliverable reminder notifications to an adolescent population.
- Areas with higher poverty rates may have distinct differences that result in higher return rates.
- Alternate sources may be a useful source for address information to improve reminder/recall activities.[53,54]

In summary, this MIIC project has been successful in improving the quality of address data and in updating patient status for those who have moved out of the area. An impressive 70% of the 11- to 12-year-old cohort was updated with new address information, and these updated addresses significantly increased the likelihood of reminders being deliverable.

IIS Example 3: Washington State: Child Profile Health Promotion Mailings

Since its inception in 1993, Washington state's Immunization Information System (IIS) has included a health promotion mailing component for parents. This component is called Child Profile and sends health promotion materials to all parents of children aged birth to six years in Washington state. Weekly uploads from the Center for Health Statistics provide birth certificate information as well as death and adoption data. The Child Profile materials

53 Kuramoto, S. Using Address Checking Services to Facilitate Reminder-Recall in Minnesota. Presented at AIRA National Conference, 2015.
54 Personal communication from Sydney Kuramoto, MIIC Informatician, February 23, 2018.

contain age-specific information about immunizations, growth, development, safety, nutrition, and other parenting issues. There are 17 mailings, timed to correspond with the American Academy of Pediatrics recommended schedule of well-child visits. Each mailing contains an age-appropriate letter and, depending on the age, may contain other useful materials. For the first 18 months, mailings are sent every 3 months. From 18 months to 6 years, mailings are sent every 6 months.[55]

An important side-benefit of these mailings is the ongoing address verification it provides. The system sends over 1.5 million pieces of mail to about 470,000 families each year. The Department of Health contracts directly with the USPS Address Correction Services (ACS) to validate and improve the quality of addresses. ACS sends daily flat files to the IIS with address corrections that are uploaded into the IIS database. The file contains the correction categories of in-state, out of state, temporary, and leftovers (addresses that do not have a forwarding address). In addition, a mailing vendor does a separate address cleansing that pulls out unmailable addresses. The IIS has developed automated reports to improve the data quality of addresses, described below.

Data Quality Reports

- **Restricted Addresses** – The IIS matches known Community Services Office (CSO) addresses against IIS patient addresses. The CSO addresses are often used for children who are in foster care or child protective services. Once matched, the address is flagged as "not valid," and Child Profile mail is not sent. The next time the child has a vaccine administered with a new address submitted, the system is updated with the current address, and mailings resume.
- **Promote Reserve Address** – If the Health Promotion Module's address and the IIS Master Record for a patient differ and a mailing has failed, the Health Promotion address is updated to have the most current address from the Master Record.
- **Unusable Names** – Obviously "fake" baby names, such as Baby Jane Doe, are considered "unusable." In these cases, the address is flagged as "not valid," and mailings are not sent. The next time patient demographics are updated, such as when birth certificate or provider data is received containing a "usable" name, the system will send out Child Profile mailings.
- **Language Switch** – This list consists of recently switched language preference from either English or Spanish, the two languages that are available for the health promotion materials. The language switch prompts a staff member to do a manual review of the patient history. This history review sometimes reveals a bad merge.
- **Long Names** – The IIS shortens or abbreviates addresses per postal acronyms, codes, or character limitations.
- **Suspicious Date of Birth** – Discrepancies that are found:
  1. Patients merged with a sibling.
  2. A vaccine or date of birth is in question and needs further investigation. For example, hep A may be listed, but they meant to put in hep B, or a birth date listed is different from the birth certificate date of birth. The clinic is contacted and a correction requested.
- **Bad Addresses** – This report, generated by the IIS, lists addresses that are only alpha characters, are less than nine characters, or contain words like bad address or homeless. Manual review of addresses is required. If the address is found not to be legitimate, then it is flagged as "not valid." The mail is turned off until the IIS receives a more current address.[56]

In summary, the Child Profile Health Promotion system provides significant benefit to IIS data accuracy. The regularity of mailings in the first six years of a child's life greatly enhances the accuracy

55 https://www.doh.wa.gov/YouandYourFamily/Immunization/ChildProfileHealthPromotion. Retrieved on February 23, 2018.

of addresses in the IIS. One indicator of success is the number of duplicate records resolved through the address scrubbing process conducted weekly before each mailing goes out. For example, in the first 11 months of 2017, Washington resolved 2,700 duplicate records through this process and saved $15,000 in mailing materials and postage.

56 Personal communications from Michelle Campbell, Washington state IIS data quality specialist, and the Child Profile team. February 23, 2018.

APPENDIX G – EXAMPLES OF REPORTS

The samples shared in this appendix offer a snapshot of the reports used by select IIS in this process at one point in time. These samples were shared for the benefit of the IIS community.

Appendix G-1. Envision Technology Partners

The data quality report shown below was developed by Envision Technology Partners and was made available to its clients to provide an overview of data quality measures. The report can be run provider by provider or at the IIS level, which summarizes all data in the IIS. A companion report is available that provides patient-specific information that allows the user to identify records that show up in these statistics for further detailed review.

Figure 2 | *Data Quality Report*

February 26, 2018

Data Quality - Statistics - Scheduled

Provider: All
Clinic: All
Date Generated: Feb 24 2018 4:12 am

PATIENT COUNTS

This section calculates the number and percentage of patients associated to the clinic according to their patient program status.

	Patients	Percentage	Percentage of Total Patients
Active Patients in Jurisdiction	2,246,204	82.32%	75.45%
Active Patients not in Jurisdiction	482,431	17.68%	16.21%
Total Active Patients	2,728,635	100.00%	91.66%
Inactive Patients - Automatically Closed by System	3,868	1.56%	0.13%
Inactive Patients - No Close Reason Indicated	244,495	98.44%	8.21%
Total Inactive Patients	248,363	100.00%	8.34%
Total Patients	2,976,998	100.00%	100.00%

PATIENT COMPLETENESS

This section calculates the number and percentage of active patient records that have a value in the following data fields.

	Patients	Percentage of Total Active Patients
Patients with First Name	2,728,635	100.00%
Patients with Last Name	2,728,635	100.00%
Patients with Middle Name	1,747,910	64.06%
Patients with DOB	2,728,157	99.98%
Patients with SSN	1,299,220	47.61%
Patients with Gender	2,723,004	99.79%
Patients with Birth State	758,930	27.81%
Patients with Birth Country	758,816	27.81%
Patients with Birth Order	331,925	12.16%
Patients with Race	1,990,623	72.95%
Patients with Ethnicity	1,946,186	71.32%
Patients with VFC Eligibility	1,424,036	52.19%
Patients < 19 with VFC Eligibility	531,342	19.47%
Patients with Mother's First Name	801,352	29.37%
Patients with Mother's Middle Name	425,830	15.61%
Patients with Mother's Maiden Name	846,968	31.04%
Patients with Mother's Last Name	770,938	28.25%

Page 1 of 4

Provider: All
Clinic: All
Date Generated: Feb 24 2018 4:12 am

Patients with 'Test' in First Name	27	0.00%
Patients with 'Test' in Last Name	380	0.01%
Patients with 'Test' in Birth Record First Name	0	0.00%
Patients with 'Test' in Birth Record Last Name	27	0.00%
Patients with 'Test' in Alias First Name	0	0.00%
Patients with 'Test' in Alias Last Name	12	0.00%
Patients with Street Name or PO Box	2,291,872	83.99%
Patients with City	2,326,254	85.25%
Patients with County	2,124,046	77.84%
Patients with State	2,330,408	85.41%
Patients with Zip Code	2,327,210	85.29%
Patients with Home Phone Number	1,896,119	69.49%
Patients with Cell Phone Number	63,936	2.34%
Patients with Email Address	131,033	4.80%
Patients with History of Varicella Disease	35,087	1.29%
Patients with Date of Varicella Disease	22,271	0.82%
Patients created by Vital Statistics Import	753,363	27.61%

VACCINATION COMPLETENESS

This section calculates the number and percentage of vaccination events (for those active patients associated to the selected clinic) that have a value in the following data fields for each category.

	All Vaccinations	Percentage of All Vaccinations	All Non-Historical Vaccinations	Percentage of All Non-Historical Vaccinations
Total Vaccinations	24,591,260	100.00%	8,576,198	100.00%
Vaccinations Added but not Administered	8,156	0.03%	7,993	0.09%
Vaccinations with Vaccination Date	24,591,260	100.00%	8,576,198	100.00%
Vaccinations with Manufacturer	11,847,676	48.18%	7,958,020	92.79%
Vaccinations with Lot Number	10,299,795	41.88%	7,981,810	93.07%
Vaccinations with Expiration Date	9,620,854	39.12%	7,767,330	90.57%
Vaccinations with NDC	2,051,041	8.34%	2,046,030	23.86%
Vaccinations with Funding Source	4,684,249	19.05%	4,575,932	53.36%
Vaccinations with Site	6,581,561	26.76%	6,212,089	72.43%
Vaccinations with Route	7,290,245	29.65%	6,742,196	78.62%
Vaccinations with Administered By	3,855,004	15.68%	3,781,029	44.09%
Non-Historical Vaccinations with VIS	4,666,726	54.41%	4,666,726	54.41%
Unspecified Formulation (UF) entered as Non-Historical Vaccines	46,148	0.54%	46,148	0.54%

Provider: All
Clinic: All
Date Generated: Feb 24 2018 4:12 am

ACCURACY, VACCINATION GIVEN AT INVALID AGE

This section calculates the number and percentage of vaccination events and the patients impacted (for patients associated to the clinic) that were given any single immunization outside of ACIP age recommendations or vaccine label guidelines for each category.

	All Vaccinations	Percentage of All Vaccinations	Number of Patients Impacted
Total Vaccinations	24,591,260	100.00%	2,652,465
DTaP / TD / TDAP Vaccinations before 6 weeks of age	8,781	0.04%	8,552
DT / DTP / DTaP Vaccinations after 7 years (2557 days) of age	36,406	0.15%	30,703
Td / Tdap Vaccinations before 7 years (2553 days) of age	9,151	0.04%	7,868
Kinrix (DTaP-IPV) Vaccinations before 4 years (1457 days) of age	3,532	0.01%	2,496
Polio Vaccinations before 6 weeks of age	13,481	0.05%	13,038
MMR Vaccinations before 1 year (361 days) of age	15,339	0.06%	14,714
Hib Vaccinations before 6 weeks of age	4,745	0.02%	4,527
Hib Vaccinations after 6 years (2190 days) of age	9,690	0.04%	6,718
Varicella (CPOX) Vaccinations before 1 year (361 days) of age	4,555	0.02%	4,421
Pneumococcal Vaccinations before 6 weeks of age	1,714	0.01%	1,660
PPV23 Vaccinations before 2 years (726 days) of age	4,582	0.02%	2,766
HPV Vaccinations before 9 years (3283 days) of age	562	0.00%	479
HPV Vaccinations after 28 years (10226 days) of age	605	0.00%	382
Cervarix Vaccinations and a gender of Male	187	0.00%	149
Hep A Vaccinations before 1 year (361 days) of age	4,472	0.02%	3,613
Meningococcal Vaccinations before 11 years of age	5,348	0.02%	4,598
Influenza Vaccinations before 6 months of age	4,824	0.02%	4,075
Rotavirus Vaccinations before 6 weeks of age	580	0.00%	563
First dose of Rotavirus after 15 weeks of age	20,106	0.08%	19,935
Rotavirus Vaccinations after 32 weeks of age	16,411	0.07%	15,457
Non-Hep B Vaccinations before 4 weeks of age	19,530	0.08%	15,130
Zoster Vaccinations before 50 years (18258 days) of age	390	0.00%	375
Vaccinations before their DOB	704	0.00%	247
Vaccinations after the Expiration Date	14,418	0.06%	12,871
Vaccinations in the future	0	0.00%	0

Provider: All
Clinic: All
Date Generated: Feb 24 2018 4:12 am

ACCURACY, VACCINATION COUNTS INCONSISTENT WITH PATIENT'S AGE

This section calculates the number and percentage of active patient records (associated to the clinic) that have a total number of vaccinations that is more than what is appropriate for their age.

	Patients	Percentage of Active Patients
Patients with more than 2 DTaP Vaccinations before 4 months of age	5,673	0.21%
Patients with more than 3 DTaP Vaccinations before 6 months of age	5,619	0.21%
Patients with more than 4 DTaP Vaccinations before 1 year of age	6,212	0.23%
Patients with more than 6 DTaP Vaccinations before 7 years of age	9,669	0.35%
Patients with more than 2 Polio Vaccinations before 4 months of age	4,133	0.15%
Patients with more than 2 HIB Vaccinations before 4 months of age	2,713	0.10%
Patients with more than 3 HIB Vaccinations before 6 months of age	8,335	0.31%
Patients with more than 4 HIB Vaccinations before 1 year of age	8,247	0.30%
Patients with more than 2 PCV Vaccinations before 4 months of age	511	0.02%
Patients with more than 3 PCV Vaccinations before 6 months of age	1,378	0.05%
Patients with more than 4 PCV Vaccinations before 1 year of age	1,505	0.06%
Patients with more than 2 HepA Vaccinations	32,163	1.18%
Patients with more than 4 HepB Vaccinations	24,830	0.91%
Patients with more than 4 HIB Vaccinations	27,996	1.03%
Patients with more than 3 HPV Vaccinations	2,382	0.09%
Patients with more than 3 Meningococcal Vaccinations	332	0.01%
Patients with more than 2 MMR Vaccinations	34,652	1.27%
Patients with more than 4 PCV Vaccinations	47,049	1.72%
Patients with more than 3 Rotavirus Vaccinations	3,278	0.12%
Patients with more than 1 Tdap Vaccinations	53,901	1.98%
Patients with more than 2 Varicella Vaccinations	14,380	0.53%
Patients with 10 or more Vaccinations on the same date	610	0.02%
Patients with 20 or more Vaccinations before 6 months of age	573	0.02%
Patients with 30 or more Vaccinations before 2 years of age	3,042	0.11%
Patients with 40 or more Vaccinations before 5 years of age	1,052	0.04%
Patients with 70 or more Vaccinations	5	0.00%

TIMELINESS

This section calculates the number and percentage of vaccinations events where the amount of time elapsed between the date of vaccination and the date it was created in the IIS is within the indicated timeframes for each category.

	All Vaccinations	Percentage of All Vaccinations
Total Vaccinations	24,591,260	100.00%
Vaccinations entered < 1 day of Administration	5,601,648	22.78%
Vaccinations entered in at least 1 day and < 2 days of Administration	1,091,468	4.44%
Vaccinations entered in at least 2 days and < 7 days of Administration	1,104,551	4.49%
Vaccinations entered in at least 7 day and < 14 days of Administration	567,879	2.31%
Vaccinations entered in at least 14 day and < 30 days of Administration	678,398	2.76%
Vaccinations entered in 30 days or more of Administration	15,547,316	63.22%

Appendix G-2. Michigan

The following report samples were taken from the Michigan Data Quality Improvement (DQI) Plan (Version 2.0). The samples show the different data quality processes an IIS can perform based on recommended frequency.

Figure 3 | *Scheduled Data Quality Processes*

Process	Schedule	Responsibility	Tool(s)
Person deduplication algorithms	Daily	MDHHS Division of Immunizations	Automated report
Vaccine deduplication algorithms	Daily	MDHHS Division of Immunizations	Automated report
72 hour submission report monitoring	Quarterly	• MDHHS Division of Immunizations • MDHHS Division of Immunizations, MCIR Regional Coordinators	MDHHS: Accreditation Report (Business Objects) MCIR Regions: Accreditation Report (Business Objects) Immunizing Provider: EXT Transfer Report
Vital records monitoring	Quarterly Annually	• Quarterly-PHBPP • Annually-MDHHS Division of Immunizations	PHBPP: Quarterly Hep B Hospital measurement reports MDHHS: Immunization Information Systems Annual Report (IISAR)
Duplicate person records	Daily	• MDHHS Division of Immunizations • MDHHS Division of Immunizations, MCIR Regional Coordinators	Regional person deduplication queue
Duplicate vaccine entries in a record	Daily	MDHHS Division of Immunizations, MCIR Regional Coordinators	Regional vaccine deduplication queue Business Objects reports
Monitoring of provider DQA reports	Ongoing pre-production	• MDHHS Division of Immunizations, MCIR Regional Coordinators • MPHI	MCIR Regions: Vaccine Deduplication Business Objects reports MPHI: IIS DQA Tool, ESSR report

Process	Schedule	Responsibility	Tool(s)
Immunization Provider review of ESSR report and correct data quality errors listed report	Daily	• MPHI • MDHHS Division of Immunizations, MCIR Regional Coordinators • MCIR end user/ Immunizing Provider	MPHI: IIS DQA Tool Report MCIR Regions & Providers: MCIR ESSR report
Review and modify ongoing policies related to IIS data quality	Annual Review	• MDHHS Division of Immunizations • MDHHS Division of Immunizations, MCIR Regional Coordinators • MDHHS Division of Immunizations, MCIR Data Quality Workgroup members	MDHHS Division of Immunizations: Review of current policies MDHHS Division of Immunizations, MCIR Regional Coordinators: Review of current procedures as they relate to current MDHHS Division of Immunization policies
Death record imports	Monthly	MDHHS Division of Immunizations	Manual monthly process from Medicaid Warehouse to IIS
Vital Record imports	Daily	MDHHS Division of Immunizations, Tech Team	Automated report
Batch assessment (updating forecast on all records in the database)	Daily	MDHHS Division of Immunizations, Tech Team	Automated report
Weekly HL7 message count	Weekly	MDHHS Division of Immunizations, Tech Team	Automated report
Sites having no weekly HL7 activity	Weekly	MDHHS Division of Immunizations, Tech Team	Automated report
First time HL7 submitters production	Weekly	MDHHS Division of Immunizations, Tech Team	Automated report
First time HL7 submitters	Weekly	MDHHS Division of Immunizations, Tech	Automated report

Figure 4 | *Newly Identified Data Quality Procedures and Reports to be Implemented*

Process	Schedule	Responsibility	Tool(s) Required
Appropriate Vaccinations by Age – A person should not have more than: • 50 vaccinations recorded in MCIR before age 5 • 35 vaccinations recorded in MCIR before 2 years of age • 70 vaccinations recorded in MCIR regardless of age.	Annually	MDHHS Division of Immunizations, MCIR Regional Coordinators	IIS design and development of business rules Business Processes defined Business Objects
A person should not have more than 6 DTaP vaccinations by age 7. Data Items: • Date of Birth • Vaccination Encounter Date • Vaccine Type	Quarterly	MDHHS Division of Immunizations, MCIR Regional Coordinators	IIS design and development of business rules Business Processes defined Business Objects
A person should not have more than 10 vaccinations per visit. Data Item: • Vaccination encounter date	Quarterly	MDHHS Division of Immunizations, MCIR Regional Coordinators	IIS design and development of business rules Business Processes defined Business Objects
Algorithms to prevent documentation of administered vaccines on DOB, excluding ACIP	Daily	MDHHS Division of Immunizations, Tech Team	MCIR Algorithm UI warning required prior to acceptance of

Process	Schedule	Responsibility	Tool(s) Required
recommended "birth" doses			any vaccine dose on date of birth, excluding Hep B
Algorithms to prevent documentation of administered vaccines post Date of Death	All vaccine encounters	MDHHS Division of Immunizations, Tech Team	MCIR Algorithm
Algorithms to prevent documentation of duplicate vaccination type administered on same day	All vaccine encounters	MDHHS Division of Immunizations, Tech Team	MCIR Algorithm
Algorithms to prevent documentation of vaccines administered post the vaccine's expiration date	All vaccine encounters for VIM users	MDHHS Division of Immunizations, Tech Team	MCIR Algorithm
Algorithms to prevent documentation of Unspecified and/or NOS vaccines unless documented as Historical doses	All vaccine encounters	MDHHS Division of Immunizations, Tech Team	MCIR Algorithm
Compliance of Submission policy for Vital Records birth certification information, i.e., EBC creation	Daily	MDHHS-Immunization Division MDHHS Division of Immunizations, Tech Team	MCIR Regional Coordinator Procedure
Allow electronic transfer of death record data, without vaccination data attached*	Ongoing	Any immunizing provider using HL7 messaging	Transfer specs HL7 guide **Requires Admission, Discharge Transfer (ADT) Messaging*
Canned IIS report of all immunizations added to a person's record: • by HL7 message • per a specified time frame	As required	MDHHS Division of Immunizations	IIS capacity/function
• including: person's first and last name, DOB, MCIR ID, Vaccine Admin Date, Vaccine Type, Lot # and Manufacturer			
Onboarding approval process	Go Live will occur within 30 days of approval	• MPHI • MDHHS Division of Immunizations, MCIR Regional Coordinators	MPHI and Regional Coordinators: • Onboarding Procedure • ESSR report
IIS queue for Potential Person deduplication, merging and approved IIS visibility	Routinely	MDHHS Division of Immunizations, MCIR Regional Coordinators	IIS Potential Person Deduplication Queue - business rules and functional specifications TBD
72-Hour Report	Routinely	Provider	Ability to generate a canned MCIR report at provider-level
Best Practice Strategies for the MCIR Provider	Referenced as required	Immunization Provider, Health Plan, Health System, and LHDs	MCIR canned reports MCIR.org Resource Library
Best Practice Strategies for Regional and MPHI MCIR Staff	Referenced as required	Regional MCIR Coordinators, MPHI	MCIR canned reports Business Objects and other AdHoc reports MCIR.org Resource Library

Appendix G-3. IIS-TIPS

The following report samples were taken from the CDC IIS-TIPS Quarterly Summary Report used for the IIS Sentinel Sites.[57] The samples show the different data quality indicators that can be measured across an IIS.

Figure 5 | *Sample from CDC IIS-TIPS Quarterly Summary Report*

Timeliness: Vaccination Record Reporting

Table 4: Timeliness of Vaccination Record Reporting — Children Born in Quarter 3, 2017

	A		Other Sites	
Time from vaccine administration to entry into IIS	**n**	**%**	**%**	**Min–Max %**
≤1 day	54,667	89	87	80–95
>1 day to ≤7 days	4,342	7	9	4–16
>7 days to ≤14 days	1,628	3	2	1–3
>14 days to ≤30 days	391	1	1	0–2
>30 days	395	1	2	0–5

Timeliness: Demographic Record Reporting

Table 5: Timeliness of Demographic Record Reporting — Children Born in Quarter 3, 2017

	A		Other Sites	
Time from birth to establishment of an IIS demographic record	**n**	**%**	**%**	**Min–Max %**
≤ 1 day	204	7	16	3–41
> 1 day to ≤ 7 days	1,028	36	42	27–60
> 7 days to ≤ 14 days	1,068	38	28	21–37
> 14 days to ≤ 30 days	335	12	8	5–12
> 30 days to ≤ 45 days	36	1	2	1–5
> 45 days	177	6	4	3–5

Completeness: Vaccinations Administered to Children Born in the Current Quarter

Table 8: Vaccinations Administered in the Current Quarter to Children Born in the Current Quarter

	Quarter 3, 2017		
Data Element	**A**	**Other Sites**	
Demographic Elements	**%**	**%**	**Min–Max %**
Sex	100	100	100–100
Race	74	83	59–96
Ethnicity	65	52	2–89
Vaccine Elements	**%**	**%**	**Min–Max %**
Vaccine manufacturer	84	84	70–93
Vaccine lot number	100	94	85–97
VFC eligibility of the vaccination	100	83	58–96

Completeness: Vaccinations Administered to Persons Aged <18 Years

Table 9: All Vaccinations Administered in the Past to All Persons Aged <18 Years

	Quarter 3, 2017		
Data Element	**A**	**Other Sites**	
Demographic Elements	**%**	**%**	**Min–Max %**
Sex	98	100	99–100
Race	78	71	50–89
Ethnicity	73	53	1–88
Vaccine Elements	**%**	**%**	**Min–Max %**
Vaccine manufacturer	66	47	36–57
Vaccine lot number	100	48	42–59
VFC eligibility of the vaccination	100	46	9–70

57 CDC. Q & A About IIS Sentinel Sites. https://www.cdc.gov/vaccines/programs/iis/activities/sentinel-sites.html.

Accuracy: Invalid Demographic Data for Persons Aged <18 Years

Table 10: Accuracy: Invalid Demographic Data for Persons Aged <18 Years — Current quarter

	Quarter 3, 2017			
Category	**A**		**Other Sites**	
	n	**%**	**%**	**Min–Max %**
Sex	0	0.00	0.00	0–0
Race	0	0.00	0.00	0–0
Ethnicity	0	0.00	0.00	0–0
Registry status (e.g. MOGE)	0	0.00	0.00	0–0
VFC eligibility	0	0.00	0.00	0–0
Demographic establishment date	5	0.00	0.00	0–0.01

"—" indicates no data submitted

Accuracy: Invalid Vaccination Data for Persons Aged <18 Years

Table 11: Accuracy: Invalid Vaccination Data for Persons Aged <18 Years — Current quarter

	Quarter 3, 2017			
Category	**A**		**Other Sites**	
	n	**%**	**%**	**Min–Max %**
Invalid MVX or MVX inconsistent with CVX	56,518	1.28	1.54	0.9–2.15
Invalid trade name code or trade name inconsistent with CVX	0	0.00	0.16	0–0.76
Invalid route or route inconsistent with CVX	0	0.00	0.29	0–0.62
Invalid administrative site or site inconsistent with route	0	0.00	2.74	0–5.27
VFC eligibility	0	0.00	0.00	0–0
Provider characteristic	0	0.00	0.00	0–0
Vaccine entry date	947	0.02	0.01	0–0.01
Record type (i.e. administered or historical)	0	0.00	0.00	0–0
Vaccine completion	0	0.00	0.00	0–0
Vaccine refusal	0	0.00	0.00	0–0

"—" indicates no data submitted

ACKNOWLEDGMENTS

The American Immunization Registry Association (AIRA) would like to acknowledge and thank the following individuals and organizations for their support and assistance with this important project:

- Individual subject matter experts (SMEs) who contributed their expertise through discussion and document review:
  - **Assiatou Bah**, *Epidemiologist*, TN IIS
  - **Brad Couse**, *Sr. Database Architect*, Envision Technology Partners
  - **Carolina Danovaro**, *Scientist*, World Health Organization
  - **Heather Crawford**, *IIS Operations and Policy Analyst*, OR IIS
  - **LaTreace Harris**, *Lead IISSB Evaluation*, Centers for Disease Control
  - **Mary Woinarowicz**, *IIS Manager*, ND IIS
  - **Michelle Campbell**, *Data Quality Specialist*, WA IIS
  - **Mike Indovina**, *Computer Scientist*, National Institute for Standards and Technology
  - **Natalie Viator**, *Informatics Analyst*, Public Health Informatics Institute
  - **Nathan Bunker**, *Sr. Technical Project Manager*, AIRA
  - **Rob Snelick**, *Project Lead*, National Institute for Standards and Technology
  - **Scott Minkin**, *Data Quality Analyst*, MA IIS
- The AIRA Assessment Steering Committee members who developed the concept, oversaw the process, and provided input at various stages of the effort.
- The AIRA board of directors who provided input at various stages of the effort and/or reviewed and provided comment on the final guide:
  - **Kim Salisbury-Keith**, *President*, Rhode Island Department of Health
  - **Amanda Harris**, *President-Elect*, Nevada Department of Health
  - **Amy Metroka**, *Vice President*, New York City Department of Health and Mental Hygiene
  - **Jenne McKibben**, *Secretary*, Oregon Immunization Program
  - **Megan Meldrum**, *Treasurer*, New York State Immunization Information System
  - Directors
    - **Aaron Bieringer**, Minnesota Immunization Information Connection
    - **Baskar Krishnamoorthy**, Florida Department of Health IIS
    - **Bridget Ahrens**, Vermont Immunization Registry
    - **Brandy Altstadter**, Scientific Technologies Corporation
    - **David McCormick**, Indiana State Department of Health
    - **Kevin Dombkowski**, University of Michigan, Child Health Evaluation and Research Unit
    - **Megan Meldrum**, New York State Immunization Information System
    - **Quan Le**, Louisiana Immunization Program

- The AIRA staff who contributed to this document's development:
  - **Alison Chi**, *AIRA Program Director*, who provided project leadership and oversight
  - **Nichole Lambrecht**, *AIRA Senior Project Manager*, who oversaw and managed the project, provided guidance and input, and edited the guide
  - **Sherry Riddick**, *Independent Public Health Consultant*, who researched, wrote, and edited the guide

- An independent copy editor, **Maureen Brody**, who reviewed and edited the guide

- Individuals in addition to the SMEs who provided feedback during the review process:
  - **Noam Arzt**, *President*, HLN Consulting
  - **Mary Beth Kurilo**, *Policy and Planning Director*, AIRA
  - **Eric Larson**, *Senior Technical Project Manager*, AIRA
  - **Marlene Lugg**, *Senior Research Scientist*, Kaiser Permanente Southern California
  - **Donna McKean**, *Vaccine Data Coordinator*, New Hampshire Division of Public Health Services
  - **Craig Newman**, *Health Research Analyst*, Northrup Grumman/IIS Support Branch, NCIRD, CDC
  - **Laura Pabst**, *Deputy Branch Chief*, IIS Support Branch, NCIRD, CDC
  - **Beth Parilla**, *Senior Program Manager*, AIRA
  - **Bea Salada**, *State MCIR Coordinator*, Michigan Department of Health and Human Services