

The Centers for Disease Control Sentinel Site Project Denominator Inflation Project White Paper

May, 2018

Participating Sites and Principal Representatives

Michigan – Rachel Potter

Minnesota – Sydney Kuramoto, Miriam Muscoplat

New York City – Alexandra Ternier

North Dakota – Dominick Fitzsimmons, Mary Woinarowicz

Oregon – Andrew Osborn

Wisconsin – Danielle Sill

Principal Editor – Andrew Osborn

Table of Contents

1. Executive Summary	Page 3
2. Introduction	Page 4
3. Evaluation Plan	Page 5
4. Record Fragmentation	Page 6
5. Moved or Gone Elsewhere (MOGE)	Page 13
6. Presentation of Collaborative Methods	Page 22
7. Conclusion	Page 24
8. Glossary	Page 26

Executive Summary

Immunization Information Systems (IIS), also known as immunization registries, are confidential, computerized, population-based systems that collect and consolidate vaccination data from vaccination providers that can be used in designing and sustaining effective immunization strategies. To be effective, IIS must maintain high data quality standards, including maintaining a consolidated client immunization record. Ideally, an IIS will contain a single demographic record for each individual within its jurisdiction. However, for reasons described in this White Paper, it is not uncommon for client record counts in an IIS to exceed jurisdictional population estimates. This phenomenon is referred to in this White Paper as Denominator Inflation (DI).

Six IIS sites, all funding recipients from the CDC Sentinel Site Project, participated in collaborative analyses to identify and measure the effect of DI in their IIS. Each site selected an issue of interest to them and designed an analysis. Another site chose to replicate the analysis and compare results. Results were shared among participant sites for review and discussion. Final written products were produced collaboratively as well. Aspects of DI that were analyzed were record fragmentation (when more than one record for a client exists in the IIS) and clients who had moved or gone elsewhere (MOGE) but were still listed as active in the IIS. Three analyses each for Record Fragmentation and MOGE were conducted and replicated by separate sites.

Comparison of results revealed that all sites could identify DI in their registries. Not all sites experience the same DI issues or to the same degree. Replication sites did not find the same magnitude of issues as originating sites. Yet with methods defined by originating sites, the effort needed to conduct the analyses by the replicating sites was vastly reduced, enabling replicating sites to explore issues that otherwise would not have been cost-effective to pursue.

The collaborative methods used in the DI project were presented at the 2017 National Meeting of the American Immunization Registry Association (AIRA). Attendees responded favorably to the broad methods and exploratory nature of collaborations.

The Denominator Project offers six analyses of DI issues for other IIS to explore and consider for usefulness in their own operational environment. More broadly, this project describes a method of inter-IIS collaboration that can easily be replicated. A library of analyses could be created that would provide a variety of benefits for all IIS, as well as for national bodies seeking to ground policies and guidelines in real world results.

2. Introduction

Immunization Information Systems (IIS), also known as immunization registries, are confidential, computerized, population-based systems that collect and consolidate vaccination data from vaccination providers, and can be used to design and sustain effective immunization strategies.¹ Initially, IIS focused on pediatric populations for several reasons: nearly all routine vaccinations initially targeted children, the vaccine schedule became increasingly complex, and children are often mobile and likely to see more than one provider within a jurisdiction. Prior to the development of IIS, clients who visited multiple providers made clinical decision-making difficult because the entire immunization history was often unknown.² IIS address this issue by creating consolidated vaccination records for every client within a jurisdiction. Over time, IIS broadened to become lifespan immunization registries and include adults. As of 2015, 51 out of 53 IIS programs in the US accepted adult vaccinations.^{3,4}

As IIS have matured, use of IIS data has expanded beyond clinical decision-making to support a number of immunization program priorities, including program evaluation. According to the Guide to Community Preventive Services, IIS are effective ways to improve vaccination rates within a jurisdiction due to their ability to assess population coverage.⁵

To be effective and reliable, IIS must maintain high data quality standards, including maintaining the consolidated client immunization records. To accomplish this, an IIS must ensure that 1) each client has a unique demographic record, to which immunization records can then be linked, and 2) records of individuals no longer residing in the jurisdiction can be identified and excluded. Ideally, an IIS will contain a single demographic record for each individual within its jurisdiction. However, for reasons such as those described below, it is not uncommon for client record counts in an IIS to exceed jurisdictional population estimates. In this White Paper, this phenomenon is referred to as Denominator Inflation (DI), named for its effect of lowering vaccination coverage estimates due to client record counts that exceed population estimates, particularly where additional client records contain incomplete immunization data. When multiple IIS records for the same client exist, DI can also impair clinical activities, such as leaving a provider unsure of which record to consult.

Beginning in 2004, the Centers for Disease Control and Prevention (CDC) Sentinel Site Project has awarded competitive funds to IIS that have achieved high participation and data quality standards and have used their IIS for program evaluation and vaccine use assessments. Currently, the CDC Sentinel Site Project funds six sites in the United States: Michigan, Minnesota, New York City, North Dakota, Oregon and Wisconsin.⁶ Even these high-functioning IIS are characterized by a count of client records that exceeds estimates for some populations. An initial analysis of the population of children four years of age and younger in each Sentinel IIS revealed a DI rate of between one and eight percent above the most recent United States Census estimates.⁷ Sentinel site representatives posed the question: what steps can be taken to identify and reduce DI within their registries, or account for it during analysis? Further, how could CDC Sentinel Site Project efforts to address DI benefit the IIS community as a whole?

Prior to this project, a site-led collaborative structure had never been undertaken for CDC Sentinel Site Project activities. The DI Project will contribute to understanding how collaborative structures can be incorporated into both general IIS and Sentinel Site Project activities.

3. Evaluation Plan

The methodology for evaluating DI consisted of two components: what issues to analyze and how to form a collaborative process among Sentinel sites to address common problems. To contain evaluation scope, the Sentinel sites agreed to evaluate demographic data only. This evaluation was generally restricted to IIS clients under age 19. For exceptions, please see the methodologies of individual analyses below.

Identifying Denominator Inflation

Two broad categories of DI were examined in this paper: Record Fragmentation and clients who have moved or gone elsewhere (MOGE). Record Fragmentation is defined as when more than one demographic record exists for a single client. This can occur when new incoming records contain data that are insufficient for automated record merging processes to work, or when they are incorrect (e.g., misspellings or the use of alternate names), and the system creates a new demographic record. MOGE is defined as unrecorded mobility of clients out of an IIS jurisdiction. All Sentinel site IIS record the immunizations of children born in their jurisdiction, and add children who have moved into their jurisdictions when those children's immunizations are submitted to an IIS. However, children also move out of a jurisdiction and providers do not always report this information to the IIS (either by inactivating a client in their system or by updating the IIS with the child's new, out-of-jurisdiction address). Thus, over time, population counts in an IIS can exceed the number of people residing in a given jurisdiction, particularly for older children.⁸

Collaborative Process Structure

All IIS jurisdictions address issues that arise in their individual IIS, and national organizations such as the American Immunization Registry Association (AIRA)⁹ provide functional guidance and best practices available to all IIS. In this evaluation, Sentinel Sites collaborated to design and replicate analyses across differing IIS architecture and operational environments. The Berkeley Collaboration Series recommends collaboration as an effective means to identify multiple techniques to solve a problem.¹⁰ The process adopted for the DI Project was intended to allow Sentinel sites to identify DI issues within their IIS and help other sites do so using replicated methodologies, but without requiring that exact techniques function in any IIS setting.

Combining Identification of DI and Collaboration Techniques

Each Sentinel site identified a known DI issue within their IIS and developed a methodology, based upon each site's internal expertise and knowledge of their system architecture and unique challenges, to specify and measure it. Each site chose also to replicate the analysis of another site that interested them. Thus, each Sentinel site became the Lead Site for one DI issue and the Replication Site for another Sentinel site's issue. Technical issues such as IIS architecture, data acceptance rules and processing procedures played a secondary role in pairing sites for collaboration. Any site could replicate any analysis that interested them. Table 1 identifies the Lead and Replication sites that worked on each analysis (see Appendix for all tables and figures).

4. Record Fragmentation

Introduction

Record Fragmentation is defined as the condition in which there is more than one record in the IIS to represent a single client. This occurs when IIS merging algorithms fail to recognize that multiple records belong to the same individual for a number of reasons: incorrect information (e.g., date of birth); the use of alternative names (e.g., nicknames) in the client record; misspellings or inverted hyphenated names; or generic names used in lieu of a client's actual name, such as fictitious names sometimes used in birthing hospitals to report vaccinations in the days before a newborn is given her or his legal name.

To ensure that client records are complete, an IIS must identify and merge all records belonging to the same individual. While all IIS use sophisticated deduplication software and processes to clean incoming and existing data, some record fragments remain. Fragmentation of IIS records results in inflated denominators and artificially depressed coverage rates at the individual and population levels. In addition, providers may over-vaccinate a patient if his or her record is fragmented, contributing to waste within the medical system. Record fragmentation in an IIS constitutes a major challenge for IIS staff and users of these data.

AIRA Modeling of Immunization Registry Operations Workgroup (MIROW) guidelines state that "Patient matching and patient deduplication are essential data processing capabilities for Immunization Information Systems (IIS)," yet acknowledge that "Currently, there is not a single idealized best practice process for patient-level deduplication."¹¹ The guidelines recommend updating and advancing deduplication technology, periodic retrospective analysis of IIS client records and the implementation of better mechanisms for sharing and collaboration among IIS regarding deduplication practices. The CDC Sentinel Site DI project offers an opportunity to form collaborative processes and develop better deduplication practices through replicated analyses of real-world record fragmentation conditions.

Minnesota, New York City and Oregon each developed a method to identify fragmented records and to assess the impact to the IIS. Their analyses were replicated, respectively, by North Dakota, Michigan and Oregon (both replicated New York City's analysis), and Wisconsin.

4a. Record Fragmentation Analysis One: Use of generic birth dates

Minnesota (lead site) and North Dakota (replication site)

Introduction

Determining accurate birth dates for Minnesota's large immigrant population is challenging. In 2015, the most common countries of origin for immigrants arriving in Minnesota were Bhutan, Burma, Democratic Republic of Congo, Ethiopia, Iraq, Somalia, and the Ukraine.¹² Many immigrants are unable to document their actual date of birth: some do not have birth certificates or are from cultures that do not celebrate births, or come from a country that uses a different calendar system.¹³ Immigrants often choose January 1 when asked to estimate their birth date because it is simple and easy to remember.¹⁴ This results in a substantial number of records with a generic birthdate. Table 2 illustrates the effect of the use of generic birth dates on Minnesota's IIS - the Minnesota Immunization Information Connection (MIIC). Within MIIC, four times as many people report being born on the first day of the year versus the subsequent two days. When attempting to merge records with a January 1 birth date, name similarities can further complicate the process. Due to these issues, MIIC does not allow automatic merging for

records that contain a January 1 birth date with the exception of records that contain evidence of birth from the Office of Vital Records (i.e., a Minnesota birth certificate number [BCN]).

Often times, immigrants and refugees have different immunization needs than those born in the United States. The lack of documentation, frequent use of generic birth dates and the need for use of catch-up schedules highlights the importance of a single, consolidated vaccination record for these clients. The goal of this analysis was to determine if other demographic data fields can be used to identify duplicate records in MIIC.

Methods

Data were extracted from MIIC for clients with a January 1 birth date, a Minnesota address and no BCN contained in the record. First and last names were extracted and combined to make a full-name variable. Birth year and gender were also extracted for validation: records would not be considered matches if the birth year and gender did not match. Descriptive variables extracted included mother's maiden name, street name, street number, city of residence and telephone number. Well populated descriptive variables from the list above were used in combination with the full-name variable and the birth year to identify possible record fragments for a single client. A sample of candidate records was visually examined to confirm possible matches.

Minnesota's replication partner for this analysis was North Dakota.

Results

A total of 62,346 records meeting the above criteria were extracted from MIIC, representing approximately one percent of all active clients. Table 3 illustrates the number of possible matches for each descriptive variable when combined with full-name. All descriptive variables except mother's maiden name, which was only populated for 7.5% of records, were retained for analysis. From retained variables, a convenience sample for each was selected for manual review. For street number and street name, the entire street address had to be an exact match. Sample records were not mutually exclusive, and could have been matched on more than one descriptive variable. Duplicates were verified by birth year and gender.

Elements of the client's address were the most effective means to identify duplicate records: full street address, street number, and street name matched 78%, 60%, and 38% of the time respectively (see Table 4).

North Dakota did not observe the same increase in the report of birth dates for January 1. From 2012-2017, counts of birth dates were as follows: January 1, 216; January 2, 201; January 3, 189. In total, there were 5,499 records in NDHS with birth dates of January 1, representing about one-half percent of all NDHS records (see Table 5.) Among identified records, street address, city and telephone number were populated at a high rate. These are required fields for NDHS record entry.

North Dakota extracted a random convenience sample of twenty-five records (approximately 4% of January 1 birth records from 2012-2017) with the variables as described above and found no matching duplicates.

Discussion

Minnesota's analysis determined that use of the street address is most effective when verifying duplicate records whose dates of birth are not reliable. Street address was more frequently populated than other fields, which may partially explain this outcome. The rate of identified matches using

descriptive data fields was lower in Minnesota than expected, perhaps due to the strict methodology for matching: all fields were analyzed using full text, and paired with a full-name variable that combined full first and last names. Methods that could account for misspellings or variants may improve results. Other fields, such as parent/guardian name, may also prove useful in identifying duplicate records,

The use of January 1 as an identifier of a generic birth date was not successful in North Dakota. North Dakota's foreign-born population is 2.7% while Minnesota's is 7.4%,¹⁵ which may partially explain the lack of comparable results. However, North Dakota has recently observed an increase in residence for populations who use alternative calendars.¹⁶ This analysis introduced the issue of generic birth dates to North Dakota staff and gave them an opportunity to test analytical processes to identify it, should it be encountered in the future.

The current analysis suggests potential for matching duplicate client records with incorrect birth dates through the use of address fields. While the success rate in this analysis was low, some records contributing to DI were identified. Replication by a Sentinel site that could not corroborate the findings indicates that the issue of generic birth dates may be specific to certain populations and not a general IIS data quality issue. This analysis also highlights the benefit of providers submitting client records with all data fields populated. Submission of complete client demographic information by providers to the IIS should continue to be encouraged.

4b. Record Fragmentation Analysis Two: Use of generic names

New York City (lead site) and Michigan and Oregon (replication sites)

Introduction

Record Fragmentation will take place when a name different from the client's legal name is used to create a record in an IIS. A common instance occurs at birthing hospitals, where there can be a time gap between the birth and the naming of an infant. In the meantime, however, immunizations administered at birth (typically a hepatitis B immunization) require a name for the immunization record to be submitted to the IIS. A birthing hospital may substitute a name such as 'Baby A' because the infant's actual name is unknown. Due to reporting standards and data exchange practices, it may not be possible to delay reporting of the immunization until the infant is named.

The acceptance of generic names in an IIS is a jurisdiction-level decision. New York City allows the submission of immunization records with generic names to their IIS, the City Immunization Registry (CIR). Michigan and Oregon maintain lists of generic names their IIS rejects, but these can be circumvented through slight modifications of future submitted names. For both Michigan and Oregon, client records with generic names exist in the IIS.

Once the record with a generic name arrives at the IIS, there is a likelihood it will not be matched to the client's authoritative record through automated deduplication algorithms; often, manual review must take place to ensure proper matching. Manual review is costly, time-consuming and prone to inconsistent decision-making. Until the fragmentation is resolved, it results in incomplete individual records in the IIS and contributes to DI. The goal of this analysis is to measure the effects of generic names on DI and to identify the most frequent generic names or name combinations that are responsible for creating duplicates in an IIS.

Methods

Data targeting two distinct cohorts were extracted: children 2-6 weeks of age (to assess record fragments created with submission of the first hepatitis B immunization), and children 19-35 months old (to assess ongoing generic duplicates cumulatively). Merging and deduplication software and processes of each participating IIS were followed independently, according to standard practices at each site. Participating sites searched sampled records manually for suspect generic names such as "Baby A" or "Newborn." Duplicates were classified as suspect records due to a generic name, due to a common first and or last name, or due to typos in the demographic records. Metrics were calculated measuring the ratio of the number of all duplicates to the number of unique children in the sample, of generic-name duplicates to the number of unique children, and the percent of duplicates accounted for by generic names.

Michigan and Oregon were the replication sites for this analysis.

Results

New York City extracted a random sample of 5,000 clients for both age groups. Michigan extracted its full population for both age groups. Oregon extracted the full population of children 2-6 weeks old (N=4,448) and a random sample of 5,000 clients aged 19-35 months.

The three sites developed a combined list and identified over 200 generic names or name combinations during this analysis. See Table 6 for a sample list. Frequently used generic names included 'Female,' 'Boy,' 'Babyboy' and 'Newborn.' Some generic names were a modification of the mother's name, such as 'Sarasgirl.'

See Table 7. In the 2-6 week age cohort, Oregon had the highest rate of duplicate records, the highest percentage of duplicate records accounted for by generic names, and the highest overall cohort inflation due to duplicate records. For the 19-35 month age cohort, New York City had the highest rate of duplicate records, the highest percentage of duplicate records accounted for by generic names, and the highest overall cohort inflation due to duplicate records. Michigan reported the lowest rate of duplicates in all cases, with the exception of being minutely higher than Oregon in cohort inflation by all duplicates for the 19-35 month age cohort. Between the 2-6 week and 19-35 month age cohorts, population inflation due to generic names dropped for all participating sites. A combined list of common generic names was developed in this process. The list included over 200 values.

Discussion

All three sites identified records with generic names in their IIS, despite both Michigan and Oregon maintaining lists of generic names to reject. Table 6 illustrates the mutability of generic names, and demonstrates the challenge of maintaining comprehensive lists of names for an IIS to identify as generic and reject. As long as data exchange protocols require immediate submission of immunization data to an IIS, a conflict between data submission requirements and elimination of the use of generic names by birthing hospitals will exist.

Michigan identified the fewest generic-named records in its IIS. Michigan employs staff whose duties include routine manual deduplication of records, while NYC and Oregon are unable to support such activities. The percent of records with generic names dropped for all three sites from the 2-6 week-old cohort to the 19-35 month cohort, suggesting that over time, records with generic names are identified and either removed from the IIS or modified to reflect names accurately and then merged with their other record fragments. This analysis illustrates that the submission of client records with generic

names contributes to DI and can even affect IIS with protocols in place to detect them. Routine analysis for generic names, and the maintenance of up-to-date lists of generic names, will assist an IIS in addressing this issue.

4c. Record Fragmentation Analysis Three: Use of numeric fields during deduplication

Oregon (lead site) and Wisconsin (replication site)

Introduction

IIS require that submissions of client records include a client's first and last names (FLN) and date of birth (DOB). Demographic records submitted to IIS in their earliest days often contained only these fields. Because of these conditions, FLN and DOB have been relied upon heavily for demographic record merging.

Internal tests conducted within Oregon's IIS – ALERT IIS – indicate that DOB is a highly reliable field, accurate for more than 99% of records.¹⁷ Challenges exist, however, with the use of FLN for record matching: names can be misspelled or changed (e.g., due to marriage or adoption), variants (such as nicknames) can be used, hyphenated last names can be inverted, or alternate names introduced (e.g., mother's maiden name on a hospital birth record). Features of matching software used by many IIS (e.g., Soundex¹⁸) have reduced record duplication due to FLN disparities. However, the nature of FLN and other alpha character fields makes them prone to error. Where algorithms can't identify a positive match with an existing record, a new client record can be created for the same individual, contributing to DI.

Other fields often exist in a client's record that can address some of the shortcomings of alpha characters, namely numeric or alpha/numeric values (such as Medicaid ID), which are often used for identification. While these fields are also vulnerable to data entry error, such errors may more likely be identified and corrected if these fields are also used for other purposes such as insurance billing or medical chart identification. Additionally, numeric or alpha/numeric fields should not be as prone to inversion or the use of variants or alternates like nicknames, potentially making them more reliable and consistent.

Numeric and alpha/numeric fields in an IIS can be well populated. For example, from 2007 – 2015, the percent of population of the telephone number field reported to the IIS Annual Report (IISAR) to the CDC averaged 64.3% (for the subset of records reported), and showed a gradual trend of increasing field population.¹⁹ An analysis of ALERT IIS data suggests that for Oregon the overall population of the telephone number field could be much higher.²⁰ National initiatives such as the American Recovery and Reinvestment Act (ARRA)²¹ and Meaningful Use²² have dramatically improved record submission and field population in an IIS, meaning that an IIS may be more likely to find numeric and alpha/numeric identification fields populated. The goal of this analysis was to determine if greater use of numeric or alpha/numeric identification fields could improve record merging processes.

Methods

Records for all clients under age nineteen in the Oregon Sentinel jurisdiction were used for this analysis (N=710,820). To test traditional FLN record matching, a variable was created using the first four letters of the last name and the first three letters of the first name (L4F3), a method frequently used in Oregon to identify candidate records for merging. Three numeric or alpha/numeric values were selected for comparison: Medicaid ID, seven-digit telephone number, and Provider Chart ID. For phone number

matches, the gender listed in the client records also had to be the same. For Chart ID comparisons, Provider Organization ID also had to match to eliminate the possibility that separate providers might use the same Chart ID number. For all tests, DOB had to match.

Demographic records that contained each of the above variables were selected. From the subset of each alpha/numeric value tested, a random sample of 250 clients was drawn for which exactly two records were included as possible matches that contained these variables. This was to ensure a clear yes/no outcome for the analysis. If fewer than 250 eligible matches were identified for a given test, all possible matches were included.

Sampled records were visually reviewed to determine whether they belonged to the same client and the results were recorded for all test variables. The order in which the numeric and alpha/numeric variables were analyzed was also noted because the same client records might be matched on multiple criteria, in which case clients merged on a previous variable would be unavailable for testing with later variables. Sampled records might also be unavailable for testing if they were merged between the time the sample was drawn and the possible matches reviewed.

Wisconsin was the replication partner for this analysis.

Results

Table 8 presents the total population of the Sentinel jurisdictions of Oregon and Wisconsin, the counts of client records containing the variables of interest, and the number of each of those records which appear to have possible matching records. There were far more duplicates identified in Oregon's IIS. Only the Medicaid ID match test in the Wisconsin replication had fewer than 250 sample records to review.

See Table 9. The percent of positive matches for records suspected of belonging to the same client was substantially higher in Oregon than in Wisconsin. In both states, a numeric value – Provider Chart ID – outperformed L4F3 in identifying record fragments. In Oregon, the Medicaid ID number also proved effective for merging at a higher rate than L4F3. For Wisconsin, there was an inadequate number of records to test Medicaid ID. Wisconsin's IIS will not allow two clients to have the same Medicaid ID, so the single match identified should be considered an anomaly. For both Oregon and Wisconsin, telephone number performed more poorly than the traditional name/DOB matching procedure.

Discussion

The use of numeric or alpha/numeric fields appears able to improve merging rates for IIS over strict use of name/DOB. However, a field being numeric doesn't ensure it will improve matching; the telephone number variable used in this analysis was less successful for merging than a client's name. Use of a field that contained a number unique to a client outperformed name/DOB merging: the Provider Chart ID proved more successful for merging than a client's name, and while Wisconsin was unable to test the Medicaid ID number effectively, Oregon results suggest it would be equally effective for merging.

Oregon's higher percentage of duplicates may be due to the fact Wisconsin employs staff whose duties include routine manual deduplication of records. But even with fewer cases to work with, the same patterns emerged in Wisconsin as in Oregon. This demonstrates that even IIS with effective merging processes may benefit from DOB/numeric-only merging tests to address challenges related to text fields in a client's record.

Other client-unique numeric or alpha/numeric variables could be tested, such as Social Security Numbers (SSN). Wisconsin does not accept SSN into their IIS in an analyzable form. Oregon captures SSN, in an encrypted but unique format suitable for comparison between records. Internal tests in Oregon have found SSN to be as effective as Medicaid ID and Provider Chart ID in identifying record fragments.²³

4d. Summary: Record Fragmentation

Analyses of record fragmentation involves the fundamentals of IIS data quality maintenance: development of methods to identify suspect records and calculations to determine the impact of an issue. The originating sites identified more records than replicating site(s), illustrating the jurisdictional uniqueness of IIS. Most sites were able to replicate a defined issue and identify at least some records affected by it, illustrating the universality of the challenges investigated. The one exception was North Dakota, who could not replicate Minnesota's generic birth date issue. However, upcoming population changes may raise the issue for them in the future.

The problems identified in this section are not the only causes for record fragmentation. Further, the methods used in these analyses are not mutually exclusive and may prove more successful if combined. For example, Minnesota's analysis assumed an incorrect date of birth was a major cause of record fragmentation, while Oregon's depended on the date of birth being accurate to identify fragmented records. New York City's analysis focused on inaccurate names. The similarities among the analyses may suggest a broader protocol that analyzes granular aspects of a client's demographic record for any of the issues focused on in this section, as well as many others.

Data acceptance now is routinely more a matter of conforming to automated data submission criteria; visual examination is becoming less and less common. Records with errors not yet detectable by acceptance algorithms can be permitted to enter the IIS. Given the speed and the volume of data submissions with which a modern IIS contends daily, it will be vital that error detection algorithms and IIS procedural updates keep pace with the rate of changes in data submission practices. In two analyses in this section, the advantages of ongoing record deduplication by dedicated staff was noted.

IIS have as a goal complete representation of a jurisdictional population. In many cases the demographic data of separate individuals can be very alike, if not identical. An IIS must determine the degree of strictness in data submission requirements and merging algorithms, and must balance the risks of record fragmentation and inadvertently merging records belonging to different people. Where to draw the line with respect to record acceptance challenges all IIS. Collaborations at the IIS level, such as those illustrated in this White Paper, offer an ideal opportunity for subject-matter experts to confer on the concrete definition of issues and methods to identify suspect records and contain record fragmentation.

5. Moved or Gone Elsewhere (MOGE)

Introduction

Often, the population in an IIS represents all people who lived or received a vaccination in the jurisdiction at any point in time, not everyone who lives there currently. The clearest evidence that an individual resides in a jurisdiction is the presence of recent accurate address data in the IIS demographic record. Address data such as street, city, state, and zip code, are core data elements in an IIS²⁴ and are often required fields to create a new record. The data are usually accurate when entered and ideally would be updated by providers when a patient presents for immunizations (which would prompt access of their IIS record). Addresses become inaccurate as patients move and do not provide updates to their providers, when providers don't update the address in the patient's Electronic Health Record (EHR) or IIS directly, or when the EHR does not transmit the update to the IIS. Functional standards state that an IIS must be able to document patient active/inactive status at both the provider site and geographic levels.²⁵

The AIRA MIROW guidance for patient active/inactive status (PAIS) was first developed in 2005 and updates in 2015.²⁶ (Note: due to the common recognition of the acronym MOGE, and because DI specifically addresses patients who have moved or gone elsewhere and not other categories of PAIS, MOGE was used throughout this White Paper.) This functionality allows providers and public health jurisdictions to indicate the status of their relationship with a particular patient and to indicate that a patient has moved away from the provider or out of the jurisdiction, permitting exclusion of the child from reminder/recall notifications and coverage assessments. Determining the correct address and a patient's relationship to the provider and local health jurisdiction is critical to effective immunization practice. Migration of patients can be difficult to determine since well-care visits decrease with increasing age, leaving the provider unaware whether the patient relocated or simply failed to make an appointment.²⁷ In addition, most people live in the state where they were born,^{28,29} so inactivation by providers may not address population inflation at the jurisdictional level.

Barring concrete confirmation through accurate address updates or system inactivation, inferences can be drawn about which records likely belong to clients who have moved out of the jurisdiction. For example, there can be a dramatic change in the pattern of a client's vaccination history. While it is possible that such a record is accurate, it's more likely the record is no longer being updated, often due to a client being MOGE. There may also be evidence of moving, such as notes by providers, in the demographic record of patients who remain active in the system but should be inactivated. To understand where the IIS may be overestimating the population, the population of an IIS can be compared to jurisdiction-level population estimates by other sources such as the US Census³⁰, which can assist in statistical adjustments to make coverage estimates more accurate.

Three Sentinel sites - Michigan, North Dakota, Wisconsin - analyzed DI caused by clients who may be MOGE. Their analyses were replicated, respectively, by Oregon, Minnesota and New York City. This section of the DI White Paper describes two methods to identify records with inaccurate addresses and one method to adjust population estimates to reflect outward mobility.

5a. MOGE Analysis One: Bad Address Data

Michigan (lead site) and Oregon (replication site)

Introduction

The AIRA MIROW guidance documents were first developed for management of clients who have patient active/inactive status and other patient designations in 2005³¹. Active/inactive status was added to Michigan's IIS, the Michigan Care Improvement Registry (MCIR) in 2011³². As described above, this functionality allows providers and public health jurisdictions to indicate the status of their relationship with a particular child and to indicate that a patient has moved away from the provider office or out of the jurisdiction, permitting exclusion of the child from recall notifications and coverage assessments. Between 1998, when MCIR was developed, and implementation of patient active/inactive status in 2011, users could flag a child's record to indicate the patient had an invalid address and eliminate the child from reminder and recall queries. Unfortunately, providers and jurisdictions have utilized neither the flag nor the MOGE functionality broadly. Determining the correct address and a patient's relationship to their provider and local health jurisdiction is critical for effective immunization practice and accurate population-level coverage estimates.

In a reminder/recall effort conducted in Michigan in 2008 and 2009, 26% of over 20,000 letters mailed were returned as undeliverable by the U.S. Postal Service. The proportion of letters returned increased with increasing age, with 11% undeliverable among children aged 6 through 18 months up to 36% undeliverable among adolescents aged 11 through 18 years.³³ A later study in Michigan compared adolescent immunization coverage using different IIS denominators and U.S. Census population estimates as a denominator. Coverage estimates for 1+ Tdap and 1+ MCV4 were 5% lower when using the IIS population that excluded inactive patients as compared to census.³⁴ Oregon has reported an 11% difference in their 1+ Tdap coverage rate among children aged 13 years, depending on whether an unadjusted IIS denominator or census estimates were used as the denominator.³⁵

Michigan has observed that medical provider staff will often use free text fields such as street address to make notes about a client's whereabouts; specifically, to note they had moved away. The goal of this analysis was to identify indicators from the street address and city fields that could suggest that a client has moved away, and flag these records for follow-up.

Michigan's replication partner for this analysis was Oregon.

Methods

Michigan identified key words and phrases in the client's street address field that might indicate a child had changed residence and whose status should be 'inactive.' Oregon compiled a similar list, and the two sites agreed on a common set of terms, shown in Table 10. Both sites determined the proportion of records in each IIS that had addresses containing these key words and compared the immunization characteristics, such as number of doses reported and time since last dose, to the records of children without these indicators. It was anticipated that demographic records with poor-quality address information, or notes questioning the client's whereabouts, would have immunization records that had not been updated for some time, suggesting that the child may have moved out of state. In addition, sites searched for telltale signs of an inaccurate address, such as there being no numbers or only numbers in a street address, or either the city or address had a null value. An accumulation of these

records in the IIS might contribute to DI and these records should be targeted for an address update and assigned the appropriate patient status, consistent with MIROW guidance. Suspect records could also then be accounted for when calculating coverage estimates.

Each site randomly selected approximately 10,000 demographic records in each of four age groups: children aged 3 through 35 months, 4 through 6 years, 11 through 12 years, and 13 through 17 years. The demographic file included street and city fields and, for Michigan, variables that have been used in the IIS to indicate patient status: county affiliation, valid address flag, jurisdictional patient status, and provider patient status. Street address fields were searched for strings, and addresses were classified as good or bad. Substrings that identified records as bad are listed in Table 10. Demographic records were linked to immunization records, and the number of doses recorded and days since last dose administered from the immunization file were counted.

Results

See Tables 11 and 12. The proportion of the sampled records with bad addresses increased with increasing age group in both sites. However, the prevalence of bad addresses in Michigan was more than double that in Oregon for the oldest age groups assessed. Very few, 0.4% in Michigan and 0.7% in Oregon, of the children aged 3 through 35 months had a bad address. Among children aged 13 through 17 years, however, almost 4% of Michigan children had bad address information compared to 1.5% for Oregon children. In both sites, the majority of bad address records contained no data in street or city fields, and the key words and strings used were more helpful to identify a bad address in Michigan than Oregon. In both sites, records with bad addresses had higher mean and median days since last immunization, and a lower number of doses reported. In addition, Michigan has several additional registry flags and indicators for use by providers to facilitate the determination of client status. As shown in Table 13, records with bad addresses were more likely to have no affiliation with a local health jurisdiction (county) and an invalid address, and were less likely to have active status at the jurisdictional or provider level.

Discussion

These results indicate that a regular review of IIS records to identify those with bad addresses would be useful for classification of patients as active or inactive, particularly among adolescents. Further review of the information contained in bad address fields is also necessary; there may be enough information to indicate that a child has moved out of state. Additional processes external to the IIS such as the use of person locator services may be needed to determine the current address of these patients. Periodic review could also indicate if this process is ongoing even after implementation of the MIROW guidance, and which providers are making notes but not using flags. Additional training for provider offices by MCIR regional staff to emphasize the correct use of patient status indicators and the benefits of their correct use, such as elimination from profile reports and reminder/recall activities, may be needed.

From a coverage estimate perspective, identification of suspect records can improve the accuracy and relevance of results. An analysis relying upon a substantial percentage of suspect records should be treated with caution, and protocols to account for suspect records should be incorporated into data preparation methods prior to analysis.

5b. MOGE Analysis Two: College Student Population Movement

North Dakota (lead site) and Minnesota (replication site)

Introduction

According to the North Dakota University System (NDUS) Spring 2016 Enrollment Survey, 49% of students enrolled in their eleven colleges and universities originated from out-of-state. This translates to approximately 3% of the state's total population. Overall, 27% of enrolled students originate from Minnesota, 16.6% come from states other than North Dakota or Minnesota, 1% comes from Canada, and 4.2% are from elsewhere outside the United States and Canada. At the time of enrollment, 73.5% of students were age 24 years and younger and 26.5% were age 25 years and older.³⁶

Students who move to North Dakota for higher education must satisfy North Dakota state immunization requirements (Figure 1). Colleges and universities also must ensure that enrolled students meet these mandates. Students may meet requirements either by providing their own immunization records or by being immunized by a healthcare provider that reports to the NDIIS. Immunizations administered to adults are not required to be submitted into NDIIS;³⁷ the information that the NDIIS captures for adults depends upon voluntary reporting by healthcare providers.

Records may appear to remain active within the NDIIS if a local health care provider reports administered or historical immunizations for students at enrollment, during their college term or post-graduate work, and the student later moves out of state. The NDIIS relies on healthcare providers to update address information for their patients, and to mark patients as inactive as appropriate. The combination of out-of-state student population size, lack of mandates for adult data entry into the NDIIS, and the high potential for retention of out-of-state records as active within the NDIIS merits analysis for its contribution to Denominator Inflation.

Although this analysis focuses on college-aged clients, college freshman can be as young as seventeen or eighteen, thus overlapping the definition of children in this White Paper. Further, as noted in the overall introduction, IIS have expanded to become lifespan systems, and an understanding of patterns in adult immunization records is becoming increasingly important. College enrollment frequently occurs at the transition from childhood to adulthood; therefore, this analysis creates a bridge for the discussion of DI into adulthood. The goal of this analysis was to seek evidence in the client record of IIS clients who may have MOGE status following entry into college.

Methods

Active records created by college or university healthcare providers were analyzed for any of the five identifiers that would indicate the student is from out of state or is no longer in the state: a non-in-state address, a non-in-state birth record, no freshman vaccinations, no vaccinations during the years in which a client would attend college (term vaccinations), and no vaccinations post-graduation. The vaccination time points were designed to allow students, typically age 18-21 years old, to enter college, complete a 4-to 5-year college degree, and to assess their vaccination status post-graduation (at 5 years). This analysis was conducted retrospectively for a 10-year period (2005 through 2015) to cover typical college and post-graduate periods. All vaccination records were considered regardless of vaccine type or validity.

The replication partner for this analysis was Minnesota. Minnesota's vaccination requirements for college students are similar to North Dakota's. Both accept exemptions from vaccination due to history of disease or for personal belief.

Results

In North Dakota, a total of 763 active records met the qualifications of potentially belonging to a student who was not a North Dakota resident and/or had left North Dakota following college enrollment (see Table 14). None of the records contained all five identifiers, and only two records contained four identifiers. Ninety-three percent (711) of the records contained at least 2 identifiers. Of the 763 records identified, 760 (99.6%), reported non-North Dakota birth state. Few suspect records reported a non-North Dakota address. Only one record indicated the client had no freshman vaccinations reported to NDHHS.

In Minnesota, where the number of out-of-state college students is comparatively smaller than in North Dakota, 462 records were identified using the same methods (Table 15). No records contained more than two identifiers. A report of no freshman vaccinations reported to MIIC was the most frequent identifier (n=439). There were a substantial number of records with a non-Minnesota address. The 'state of birth' field in MIIC is populated from vital records loads and is therefore not populated for those born outside of Minnesota. A substantial portion of records were missing state and county identifiers for address.

Discussion

In North Dakota, almost all records indicated that freshmen vaccinations were received by the cohort. This suggests that schools and freshmen are complying strongly with North Dakota college-entry vaccination requirements. Due to the high population of the 'birth state' field, North Dakota was able to identify college students who were not born in the state, had received their freshman vaccinations, but then did not receive any immunizations during either the mid-college or post-graduation timeframe. These records are likely those for clients who attended school in North Dakota but then left either before or upon graduation. Few suspect records reported a non-North Dakota address. While attending college, students can adopt a local address even if they don't intend to remain in the state.

Despite the large number of out-of-state students in North Dakota relative to population, this group contributed fewer numbers than expected to DI. However, for the college age group, the out-of-state birth status and an abrupt change in vaccination pattern were strong indications of records that were not complete, most likely due to the student leaving the state. In that respect, this analysis likely identified one source of DI in the NDHHS. In cases such as this, further corroborating evidence should be obtained, such as address confirmation by providers or address updates from alumni associations.

Results in Minnesota differed substantially. In a state that is more than seven times larger than North Dakota in terms of population,³⁸ fewer than two-thirds of the number of suspect records were identified. Fewer out-of-state students may attend Minnesota schools, or they may establish residency there during or after college, and thus their records become those of residents. Lack of population of the 'state of birth' field in MIIC prevented Minnesota from analyzing the effect of out-of-state birth on vaccination patterns. Differences in immunization requirements for college entry may also have affected the results.

Under the assumption that all higher education institutions with residence facilities receive students from out-of-jurisdiction, the migration of the college-aged population in and out of a given jurisdiction can lead to the addition of records to an IIS for young adults who are no longer there. Over time, this can produce an incremental influence on DI. In North Dakota, the number of records possibly contributing to DI represent approximately 3% of the state's population between the ages 25-29 – substantial enough to influence coverage rates. As interest in the immunization records of the adult population grows, identification of young adults who no longer live in a given jurisdiction, yet whose IIS record appears to be active, will become more important. Immunization records of young adults in an IIS that are not updated could contribute to DI for decades. This first effort to identify such records may allow further analyses to build upon these ideas and premises.

5c. MOGE Analysis Three: Accounting for Out-Migration in Population Estimation

Wisconsin (lead site) and New York City (replication site)

Introduction

When an IIS cannot confirm that a client has moved away, the focus of addressing DI can shift from client inactivation to that of accounting for potentially inactive records during analysis. According to U.S. Census estimates, 5.7 million people resided in Wisconsin (WI) in 2015.³⁹ At the same time, 6.5 million clients were categorized as WI residents within the Wisconsin Immunization Registry (WIR).⁴⁰ A potential contributor to population overestimation in WIR is the group of clients who move out-of-state but remain active in the IIS. WI uses all active WIR client records to calculate coverage rates and identify pockets of need. Denominator Inflation – especially in younger age groups, when most vaccination takes place – makes the calculation of accurate vaccination rates challenging. This analysis focused on determining if migration out of an IIS jurisdiction (i.e., out-migration) could be measured by comparing population estimates from different sources (i.e., the IIS and Census), with the goal of increasing the accuracy of vaccination coverage calculations.

Wisconsin's replication partner for this analysis was New York City (NYC).

Methods

WI used state-to-state migration flow tables created by the U.S. Census Bureau in the American Community Survey (ACS)⁴¹ to determine out-migration from WI. The ACS is a national survey collected monthly from individuals who completed the decennial long-form census. In WI, between 61,000 and 75,000 individuals participate in the ACS survey annually.⁴² The ACS was used to determine the total number of individuals who moved out of WI each year by calculating the number of persons who reported their state of residence as WI a year prior and currently resided in another U.S. state or territory. Data for WI were available for ten years, from 2005 through 2014. However, age group migration counts were not available in state-to-state migration tables. For NYC, state-to-state migration tables could not be used to estimate migration as the New York State population is not representative of that of NYC. Instead, the Integrated Public Use Microdata Series dataset (IPUMS) was used to obtain NYC metropolitan-level migration estimates.⁴³ The IPUMS has collected data from over fifty survey samples conducted by the U.S. Census or ACS. NYC was able to retrieve metropolitan migration data for a seven-year time period, from 2005-2011 (city-level migration estimates for NYC were unavailable from

2012-2014). In addition, we used regional out-migration estimates from the Midwest and Northeast available from the U.S. Census Bureau's Annual Social and Economic Supplement. These estimates were stratified by age group.⁴⁴ Age-stratified state-to-state or IPUMS regional migration percentages were multiplied by the population counts to calculate the number of individuals who moved out of jurisdiction in the 1-19 year old age group. This total was subtracted from the IIS population to create the Expected IIS Population, representing the IIS population if out-migrants were removed. Using the estimated number of individuals who migrated out of the jurisdiction per year, we calculated the Compounded Migration by summing out-migration counts cumulatively over the entire time period studied for each jurisdiction (seven years for NYC and ten years for WI).

Results

As shown in Table 16, Compounded Migration estimates were 296,640 for WI and 60,351 for NYC during the ten and seven year time periods, respectively. Between 2005 and 2011, WI saw a total of 213,630 out-migrants depart their jurisdiction compared to 60,351 in NYC. The highest out-migration counts were 33,097 persons in WI in 2008 and 14,944 persons in NYC in 2011. During the study period, we found fluctuations in the annual migration counts for WI. In contrast, the NYC estimates showed a steady upward trend for the period examined. Overall, WI had higher annual out-migration counts compared to NYC.

The difference between the census and IIS population counts attributable to out-migration is shown in Table 17. Out-migration accounted for 16.3% to 73.1% of the difference between the IIS and Census population estimates in WI and 1.0% to 7.9% in NYC. Over the study period, the difference attributable to out-migration, on average, decreased for WI but remained relatively stable for NYC, with the exception of the year 2005.

Discussion

IIS population counts can deviate significantly from actual population counts in the absence of a way to identify individuals who migrate in or out of the IIS jurisdiction. This analysis showed the impact of out-migration on population counts in WI and NYC, where a significant number of former residents remain active in the respective IIS. Between 2005 and 2011, compounded out-migration was significantly higher in WI compared to NYC.

In 2005, the WIR was populated with fewer individuals than the census estimate. This difference could be attributed to organizations that were not using the WIR in 2005, but started sending immunization data to the WIR in 2006 or later.

Out-migration in each jurisdiction fluctuated from year to year, as did the total difference between the IIS population and the census population. Even with a larger population and smaller geographical area compared to WI, migration remained lower in NYC. It may be that regional age estimates used to calculate out-migration were not representative of the NYC area, whereas WI aligned more closely with their comparison estimates. Contemporaneous events could also explain some of the observed patterns. For example, the difference between the IIS and census population was highest in NYC in 2005, which may be due to 9/11-related (residual) out-migration. Differences in system architecture, data processing and cleaning procedures, and other factors may also help explain the differing results. For example, we did not account for record duplication (i.e., multiple records belonging to the same individual), which could have affected the population differences calculated in this analysis.

In WI, the expected IIS population estimate grew closer to census estimates when out-migrants were removed from each age category. As was seen in NYC over the entire time period, if IIS populations are far from census estimates, removing out-migrants may not dramatically improve the expected IIS population. When working with jurisdiction-level migration estimates, data are not available until almost a year after the data are received, making timely adjustment for out-migration a challenge. An overall out-migration rate was calculated in this analysis using available data sources. However, these estimates may be biased by factors such as granularity of available data, stability of an age group (even within a year), the years used for analysis, and jurisdiction definition.

While this analysis met with challenges, it begins to address an important IIS issue. Removing out-migrants and remaining aware of out-migration can help reduce DI if the IIS population is close to the census estimate. Over time, analytical methods could be refined as migration patterns become better defined and additional reliable data sources for comparing IIS population counts are identified. Further, these methods could be combined with other evidence (such as length of record inactivity) or with other statistical methods such as data weighting to estimate accurate population denominators. Monitoring and adjusting for migration patterns can help reduce DI and improve the accuracy of coverage estimates in an IIS.

5d. Summary: MOGE

The analyses in the MOGE section of this White Paper were substantially broader than those employed to address Record Fragmentation, moving from a precise, record-specific analysis looking for bad addresses, to an analysis seeking to measure the movement of a specified population, to an effort to address migration a jurisdictional level. Even the most precise methods here required a greater degree of inference to be effective. An important issue, then, for the IIS community will be to define how much inference is acceptable: how confident can one be in the conclusion drawn from this type of analysis? As with Record Fragmentation analyses, the originating site of the analysis observed a greater degree of the issue, yet the issues were also identified in the replication sites. Also as with Record Fragmentation, there are more issues with clients who may be MOGE than were analyzed in this section.

The methods used in these analyses, while conducted independently, might be most effective if combined. Combining record-level information, analysis of the movements of specific populations, and jurisdictional-level approaches could improve the effectiveness of overall record identification. Such a process could create a more precise denominator for coverage estimates. Further combination with existing data weighting methods would likely improve the accuracy of estimates even more.

The MOGE analyses introduced new levels of IIS uniqueness beyond architecture and functionality: that of population environment. Use of the analyses in this section by an IIS in a jurisdiction with a stable population would look very different from one where population composition is changing rapidly, even if all other variables are the same. As IIS continue to develop and the IIS community becomes more tightly interwoven through practices like data exchange, a better understanding of population movement at all levels will be necessary.

With the expansion of IIS activity into adulthood, the need to account for clients who have MOGE status will increase. A record of a client in a childhood age group who has moved away could deflate coverage estimates for up to several years. By contrast, reports of immunization coverage for adults can use age

groups that are several decades in range, compounding the effects of DI on a given estimate. To meet this challenge, the IIS community will need to make use of every tool at hand.

6. Presentation of Collaborative Methods

Introduction

A key component of the Denominator Inflation project was the collaborative nature of work among Sentinel sites. The group collaborated on the general topic of DI and individual analytical ideas. Sites chose to replicate a given analysis based upon interest in the topic. Analyses were sometimes modified by replication sites out of necessity (being unable to replicate precisely the methods of the originating site), expedience (e.g., challenges in exact replication extending beyond the scope of the overall project) or inspiration (improving the initial analytical proposal). All Sentinel sites were afforded the opportunity to replicate any analysis of interest, thus the potential existed for multiple layers of analytical modification. In this pilot project for collaboration, important issues in IIS operations such as IIS architecture, data acceptance rules, what data fields are stored in an IIS, and comparability of jurisdictional populations were not considered. Sentinel sites wished to present the collaborative methods used to the IIS community, to share experiences and gain reactions to the DI project as it was designed and as it developed.

Methods

In April 2017, representatives from the Sentinel Sites presented the collaborative framework of the DI Project as part of the 2017 AIRA National Meeting in Chicago, IL.⁴⁵ After an introduction to the project and a description of methods, audience members were broken into four sub-groups, each led by a Sentinel Site representative, to share their thoughts on participating in such a collaborative effort. Sub-group participants were encouraged to think expansively about the DI collaborative model, their own collaborative practices, barriers to further inter-IIS collaborations, and suggestions for development of the concept. Three areas for discussion were proposed to the sub-groups: whether to focus on the rigor of a proposed analysis or to encourage approximate replications as well, should IIS of similar or different architecture collaborate, and should analytical partners remain the original participants throughout or could new partners join during the process. Participants were asked to select along a continuum of choices they would recommend.

Results

Presentation attendees indicated support for a more open, community-focused collaborative process, but without neglecting the need for adequate rigor or the opportunity for initial sites to ensure that an issue was well defined and described before inviting additional sites to participate. The need for IIS to learn from each other, and for more and broader forums for collaboration among IIS, was voiced strongly. Attendees also made it clear that usable results must be obtained from collaborations in order to justify the work. Whether defined tightly or loosely, attendees saw benefits in greater collaborations, and the chance to make use of other IIS collaborative findings.

Discussion

Based upon audience/participant feedback, the broad and variable nature of the DI project appeared to be on target with what other IIS staff feel would work generally. The AIRA National Meeting is the primary community meeting for IIS staff, and other presentations were taking place at the same time; attendees of the DI Collaborative presentation may have self-selected as being biased toward the use of open, community-level collaborations. Also, as noted in the introduction, the DI collaborative model is unique, so the novelty of the project plan may also have biased audience members favorably.

Nevertheless, attendee responses to the presentation indicate that the collaborative structure of the DI project has appeal. Future collaborative efforts of this kind should provide adequate structure without becoming overly stringent with regard to analysis-level protocols and defined processes, definitions and parameters. The CDC 2013-2017 IIS Functional Standards note the importance of expert rule development (p39). If the support for IIS-IIS analytical collaboration extends to the broader community, projects like the CDC Sentinel DI project could make an important contribution to the improvement of quality in all IIS.

7. Conclusion: Denominator Inflation Project

Since their inception, IIS have emerged as a key component of electronic health data systems. They are intertwined with medical offices, hospitals and local health departments regarding vaccination coverage, emergency preparedness, reminder/recall, as well as a host of other activities. The effectiveness of IIS is reliant upon the quality of data stored in them. Denominator Inflation affects even the best IIS and is a constant challenge. Addressing DI successfully requires vigilance, and the effort will be aided through the types of analyses and collaborative methods described and applied in this White Paper.

The conclusions and applications of the CDC Sentinel Site Denominator Inflation Project are manifold. Taken at a granular level, each participating IIS identified a condition that increased DI in its IIS, and by identifying it paved the way for addressing it in the IIS or accounting for it during data analysis. Using a relatively simple analysis, every high-quality Sentinel Site IIS identified a DI issue and in almost every case, another IIS independently identified the same issue, even if not to the same degree and magnitude. Denominator Inflation is a universal issue for IIS, but environmental conditions can alter the degree to which a given issue poses a substantial problem due to factors such as population composition, local laws and policies, and IIS business rules and architecture.

At a broader level, the combined analyses for both Record Fragmentation and clients who are MOGE addressed categorical issues in ways that can be used independently, combined for greater possible effect, or as an inspiration and springboard to additional methods of analysis. For Sentinel Sites that explored Record Fragmentation, findings included that for populations that use non-standard calendars, generic birth dates can increase DI. The use of generic names, particularly by birthing hospitals, can increase the odds that portions of a client's immunization record are not identified and merged. The use of numeric and alpha/numeric data fields in a client's demographic record can increase the likelihood that record fragments are identified and united. In analyses examining clients who have MOGE status, there were several key findings. Bad addresses offer a clue to identifying clients who may have moved away. The transient nature of the college-aged population can contribute to DI, particularly in those states where college attendance by out-of-state students is high. Independent population estimates can be used to adjust the denominators from an IIS, and thus improve the accuracy of coverage estimates.

There were several limitations associated with this project. The list of DI issues explored was not exhaustive or comprehensive. The projects analyzed and replicated in this study were specific issues identified by each Sentinel Site. Replications were not always able to implement exact methodologies. This project focused on identifying DI, but did not make recommendations to handle it in an IIS.

As a collaborative model, the DI Project responds to the call from MIROW to increase cooperation and idea exchange among IIS. The analytical methods used for the DI Project were relatively simple, easy to conduct (if somewhat imprecisely), and effective in identifying individual IIS data quality issues. In many respects, these analyses resemble data quality methods used in an IIS as part of routine daily operations, with the exception that they were described and replicated as closely as possible by another IIS. The authors of this White Paper speculate that there are dozens, if not hundreds, of analyses within IIS community that could have fit within the criteria of this White Paper. We hope that other IIS will be eager to review our analyses for applicability to their own registries. Indeed, a natural extension of the DI project would be to see how many DI issues each Sentinel Site could replicate. What is needed to facilitate such a rich information exchange are a venue and a means. We propose that national bodies consider the creation of a library of IIS data quality analyses, to be accessed by all IIS for reference, replication and expansion. Using straightforward analyses such as those in the DI Project, the work of any IIS could be stored and indexed for use by all registries. At the national level, governing bodies

could use library contents to inform policies and recommendations that would be based on concrete, demonstrable outcomes from the real world, as well as develop shared standards and practices for data quality analysis.

At the IIS-level, such a library could save enormous time devising methods to identify data quality issues, and have the added benefit of another registry's results for use as a comparison. In this White Paper, replication sites always found a defined issue to have a smaller impact than the originating site. One positive outcome of this project, then, was to permit Sentinel Site IIS to minimize the amount of effort needed to analyze an issue of unknown size. A library could dramatically alter the cost/benefit calculus of exploring unknown IIS data quality issues. An extended local benefit of a library would be its use as a training tool for new staff, both to familiarize them with data quality methodologies, and to enhance collaboration and writing techniques. The presentation at the 2017 AIRA National Meeting suggests that a substantial portion of the IIS community would welcome the opportunity to collaborate on a library of analyses.

The work contained in this White Paper should serve as a call to the IIS community to pursue DI collaboratively and improve and expand upon this work. Such work could include new analyses, confirmation of existing analysis, and concrete steps for any IIS to take to address some of the causes of DI in their IIS.

Glossary of Acronyms

Acronyms are listed alphabetically, not in the order they appear.

ACS – American Community Survey

AIRA – American Immunization Registry Association

ALERT IIS – Oregon Immunization Information System

ARRA – American Recovery and Reinvestment Act

CDC – Centers for Disease Control and Prevention

CIR – City Immunization Registry (New York City’s IIS)

DI – Denominator Inflation

EHR – Electronic Health Record

IIS – Immunization Information System

IPUMS – Integrated Public Use Microdata Series

L4F3 – First four letters of the last name, first three letters of the first name (Oregon analysis)

MICR – Michigan Care Improvement Registry

MCV4 – Meningococcal Vaccine

MIIC – Minnesota Immunization Information Connection

MIROW – AIRA Modeling of Immunization Registry Operations Workgroup

MOGE – Moved or Gone Elsewhere

NDIIS – North Dakota Immunization Information System

NDUS – North Dakota University System

NYC – New York City

PAIS – Patient active, inactive status

SSN – Social Security Number

Tdap – Adult Tetanus, Diphtheria, Pertussis Vaccine

WIR – Wisconsin Immunization Registry

References

- ¹ Centers for Disease Control and Prevention (CDC). Progress in Immunization Information Systems- United States, 2012. MMWR 2013; 62(49):1005-1008. Retrieved at: <https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6249a4.htm>
- ² Smith PJ, Stevenson J. Racial/ethnic disparities in vaccination coverage by 19 months of age: an evaluation of the impact of missing data resulting from record scattering. Stat Med 2008;27:4107–18.
- ³ National Vaccine Advisory Committee. Recommendations from the National Vaccine Advisory Committee: Standards for Adult Immunization Practice. Public Health Reports. 2014;129(2):115-123. Retrieved at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3904889/>
- ⁴ Martin DW, Lowery NE, Brand B, Gold R, Horlick G. Immunization Information Systems: A Decade of Progress in Law and Policy. J Public Health Manag Pract. 2015;21(3): 296-303.
- ⁵ Groom H, Hopkins D, Pabst L, et al. Immunization Information Systems to Increase Vaccination Rates: A Community Guide Systematic Review. J Public Health Manag Pract 2015; 21(3): 227-248.
- ⁶ <https://www.cdc.gov/vaccines/programs/iis/activities/sentinel-sites.html>
- ⁷ Unpublished data, CDC Sentinel Site Project, September, 2014
- ⁸ Robison, Steve. (2015) Addressing immunization registry population inflation in adolescent immunization rates. Public Health Reports, Mar-Apr, 130(2), 161-166
- ⁹ <http://www.immregistries.org/>
- ¹⁰ Stadelman L. Successful Collaboration: An Overview. [Brochure]. Toronto, Ontario, Canada: Berkeley Consulting Group. Retrieved at: http://www.berkeleyconsulting.com/strategic/Successful%20Collaboration_An%20Overview.pdf
- ¹¹ <https://www.cdc.gov/vaccines/programs/iis/interop-proj/downloads/de-duplication.pdf>
- ¹² <https://www.acf.hhs.gov/orr/resource/fy-2015-refugees-by-state-and-country-of-origin-all-served-populations>
- ¹³ <http://www.mncompass.org/immigration/overview>
- ¹⁴ <https://www.mprnews.org/story/2009/12/29/january-1-birthdays>
- ¹⁵ <https://www.americanimmigrationcouncil.org/topics/state-by-state>
- ¹⁶ Personal correspondence, Dominick Fitzsimmons, NDIIS, January 29, 2018
- ¹⁷ Unpublished data, Oregon ALERT IIS, September, 2016
- ¹⁸ <https://en.wikipedia.org/wiki/Soundex>
- ¹⁹ Personal correspondence, Terence Ng, CDC, May, 2017
- ²⁰ Unpublished data, Oregon ALERT IIS, May, 2017
- ²¹ <http://www.ffis.org/452501/453806.html>
- ²² <https://www.healthit.gov/providers-professionals/meaningful-use-definition-objectives>
- ²³ Unpublished data, Oregon ALERT IIS, Sept, 2016
- ²⁴ 2013-2017 Immunization Information System (IIS) Functional Standards. <https://www.cdc.gov/vaccines/programs/iis/functional-standards/func-stds-2013-2017.html>
- ²⁵ <https://www.cdc.gov/vaccines/programs/iis/functional-standards/func-stds-2013-2017.html>
- ²⁶ http://repository.immregistries.org/files/resources/5835adc2dad8d/mirow_pais_full_guide.pdf
- ²⁷ Gowda C, Dong S, Potter RC, et al. A systematic evaluation of different methods for calculating adolescent vaccination levels using immunization information system data. Public Health Reports (2013); 128: 489-497.

-
- ²⁸ <https://www.census.gov/prod/2011pubs/acsbr10-07.pdf>
- ²⁹ <https://www.census.gov/content/dam/Census/library/publications/2012/demo/p20-567.pdf>
- ³⁰ <https://www.census.gov/data.html>
- ³¹ AIRA Modeling of Immunization Registry Operations Work Group (eds). Management of Moved or Gone Elsewhere (MOGE) Status and Other Patient Designations in Immunization Information Systems. Atlanta, GA: American Immunization Registry Association. December 2005. http://www.immregistries.org/resources/MIROW-MOGE_Chapter_Final_122005_rev1.doc
- ³² MCIR Patient Status for Providers / Clinics. Michigan Department of Community Health. May 2011. https://www.mcir.org/wp-content/uploads/2014/08/Patient_Status_for_Providers.pdf
- ³³ Dombkowski DJ, Reeves SL, Dong S, et al. Assessing the burden of undeliverable immunization reminder and recall notifications. *Preventive Medicine* (2011); 53: 424-426.
- ³⁴ Gowda C, Dong S, Potter RC, et al. A systematic evaluation of different methods for calculating adolescent vaccination levels using immunization information system data. *Public Health Reports* (2013); 128: 489-497.
- ³⁵ Robison SG. Addressing Immunization Registry Population Inflation in Adolescent Immunization Rates. *Public Health Reports* (2015); 130: 161-166.
- ³⁶ North Dakota University System, *2016 Spring Enrollment Report*, February 2016
- ³⁷ North Dakota Century Code, 23-01-05.3
- ³⁸ United States Census Bureau, Population Division, Annual Estimates of the Resident Population for the United States, Regions, States, and Puerto Rico: April 1, 2010 to July 1, 2016 (NST-EST2016-01), December, 2016
- ³⁹ https://factfinder.census.gov/faces/nav/jsf/pages/community_facts.xhtml
- ⁴⁰ Unpublished data, Wisconsin Immunization Registry, August 2017.
- ⁴¹ <https://www.census.gov/data/tables/time-series/demo/geographic-mobility/state-to-state-migration.html>
- ⁴² <https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/sample-size/index.php>
- ⁴³ Steven Ruggles, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. Integrated Public Use Microdata Series: Version 7.0 [dataset]. Minneapolis, MN: University of Minnesota, 2017. <https://doi.org/10.18128/D010.V7.0>
- ⁴⁴ <https://www.census.gov/topics/population/migration/data/tables.html>
- ⁴⁵ <http://www.immregistries.org/events/2017/04/11/aira-2017-iis-national-meeting>