



Topic: Removing Imprecise Records

Request Date: October 10, 2018

Information Requested:

We are wanting to remove records from our system that are marked baby boy, baby girl, baby, boy, B1, BB1, BG1, twin1, twinA, etc, and would like some insight, criteria and any potential best practices for removing imprecise records from the IIS. On the side of caution, maybe a cutoff would be records that have reached 3-5+ yrs. old that have less than 2 vaccines. We are wanting more accurate data and a way to clean up anything duplicated without compromising any record that may be valid.

Requesting Member: Jeffrey McIntyre (MS)

Responding Member(s): Andrew Osborn (OR), Cassandra Ott (San Diego)

Results:

OR:

Attached a CDC Sentinel Site White Paper from earlier this year discussing 'Denominator Inflation,' the phenomenon of increasing client records, in part as a result of this problem. The White Paper is the culmination of a multi-year investigation into such issues by five states and New York City – all participants in the CDC Sentinel Site Project. In addition to the baby names analysis, there are five additional analyses in the paper that address other accuracy challenges in demographic records.



Denominator
Inflation White Paper



Denominator
Inflation Appendix F

San Diego:

We're currently addressing this issue within the San Diego Immunization Registry (SDIR). In a sample dataset of children 19-35 months old, records were flagged if they contained "baby", "boy", or "girl" anywhere in the first or last name field (case in-sensitive), with the exception that the last name must exactly equal "boy" ("boy" is a common component of last names). This accommodates variations of "newborn baby", "baby girl", "newborn boy", etc. We haven't found any other key words for this type of record, although those would



AIRA
AMERICAN IMMUNIZATION
REGISTRY ASSOCIATION

INFORMATION REQUEST

also need to be included. Then, we manually looked through a sample of these records to identify patterns by provider source. We found that the majority of the records came from interface sources with one of two naming conventions: 1) the child's first name contained "BOY[or GIRL]-mother's first name", or 2) the child's first name contained "BOY[GIRL]" and the child's middle name contained the mother's first name. For these sources, the mother's name fields were blank. As a solution, we're adjusting the match-merge algorithm to recognize the mother's name in the child's name fields, making it more likely to match these types of records.

We don't intend to delete these records because they hold valuable information. Rather, we intend to flag them, making them easily excluded from analysis/reports but still available for match-merge and provider use. However, we don't necessarily need to flag all records with a "baby" name. If a "baby" record has been merged with another record with a full name, then the record is no longer a questionable "baby" record and does not need to be flagged. Instead, the child's active/inactive status can be evaluated based on other standard requirements.

A cut-off based on time and <2 vaccines could feasibly be used to flag problematic "baby" records as well, but we think it's unnecessary for SDIR. In San Diego County, we expect nearly all "baby" records to be un-matched duplicates to another record in the database because SDIR routinely receives birth records from the Office of Vital Records. If we only flag "baby" records >3 years old with <2 vaccines, we will probably allow duplicate records in analysis/reports. Instead, we want to flag all unlinked "baby" records. However, if you can't assume most are unmatched duplicates then this may be a good rule. If it's used, we recommend only applying this rule to un-linked records. We're interested to learn other opinions on using time and number of vaccines to flag "baby" records.