NACOR Data Validation

Unlike similar medical registries NACOR takes any anesthesia related data from any source without any human intervention. The primary benefit of this method of data capture used by NACOR is that it allows for the speedy submission of much larger volumes of case data. The potential weakness of this capture method is that there may be inconsistences in the data or erroneous data values. AQI has implemented a series of validation routines to mitigate the risk of storing and presenting bad data.

Validation Routines

NACOR's data validation routines come in two forms. One set determines the correctness of incoming data prior to insertion. The other set verifies the integrity of all NACOR cases. This overlap in validation creates a tight net that ensures all NACOR case data is accurate.

Incoming data validation

The first data integrity checkpoint comes at the point of data insertion into staging tables. Incoming data is thoroughly scrubbed to ensure precision. Every piece of information is checked to be valid and within a range consistent with that particular data source.

The first data integrity checkpoint comes at the point of data insertion. Before data is inserted into NACOR, AQI personnel preform a series of checks against the data looking for potential errors. Every piece of information is checked against pre-determined value ranges.

CPT or patient demographic information such as age is a good example. Is the age range reported between 0 and 110? How are the ages distributed? If the reported range is incomplete (only adult ages for example) has this source of data always been this way? Does a single patient have more than one age? Every variable captured undergoes similar checks and all of those automated tests must pass satisfactorily before data is permitted to become part of NACOR.

Post insertion auditing

NACOR runs a daily automated job to validate all records in NACOR. These checks serve to ensure maximum data integrity. Deeper probing can be done in the form of trend analysis, comparing historically provided information with newer data submissions using daily reports generated each night.

Examining cases over time illustrates this idea well. Do the cases have an equal distribution over time over multiple data submissions? Do cases wax and wane based on the day of the week (there are less cases performed on the weekend)? If duplicate cases were inserted into NACOR a look at cases over time would easily spot this error. This data would be flagged as suspect and subsequently removed.

NACOR Referential Integrity

NACOR is a relational database with a strong referential integrity to ensure consistency of anesthesia data. This means that every record in NACOR contains values that are well defined (reference values), for example, Anesthesia Type, Gender, and Timing Events.

Validation completion

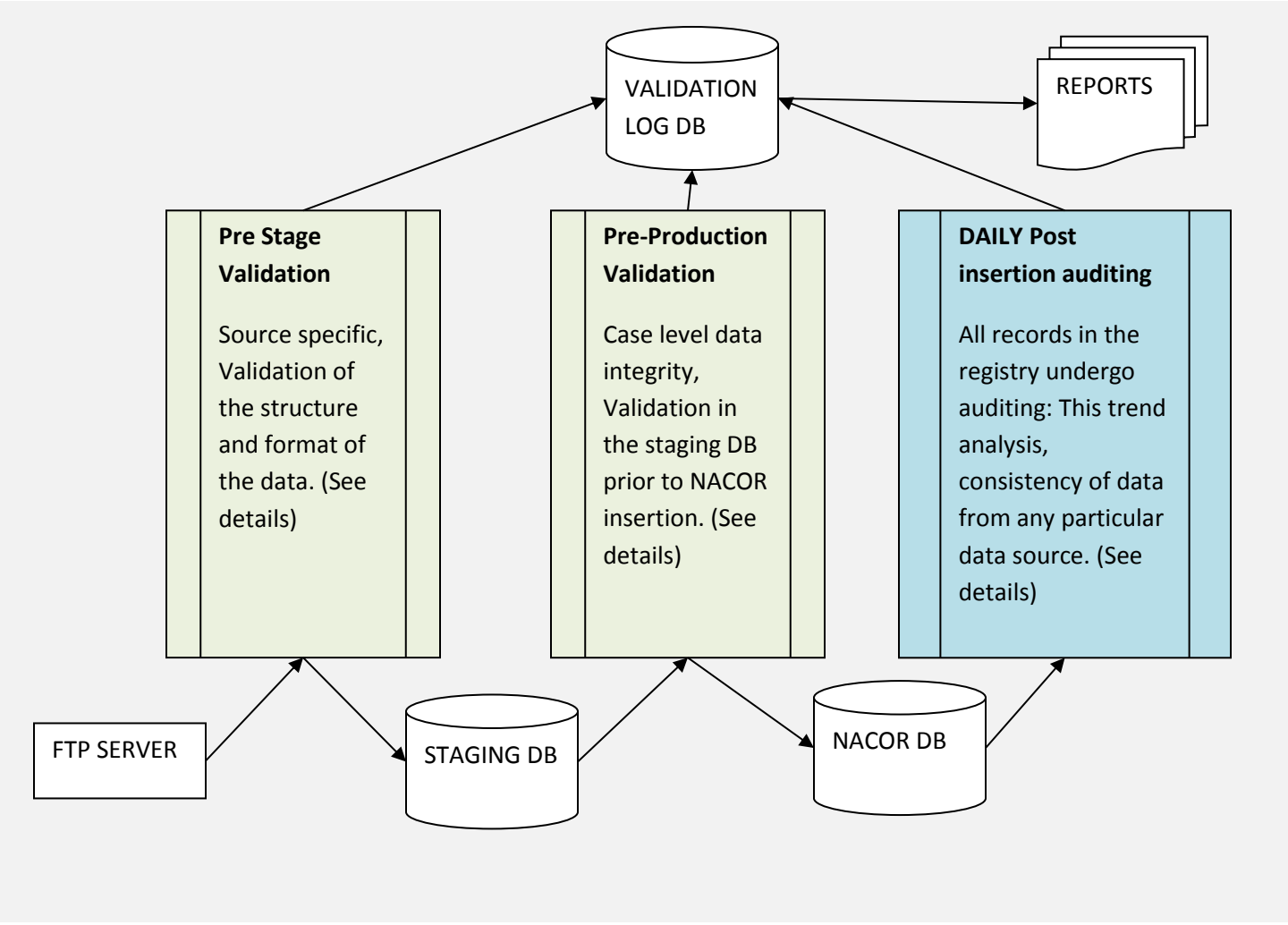
All case data is subject to intense scrutiny upon entry into NACOR. This is supplemented with a continual, more focused authentication process that last the entire lifetime of the data within the registry. With a two tiered validation method NACOR retains its accuracy while maintaining its ability to absorb large amounts of information.

Validation Details

Validation Stages

- Pre-Stage Validation
- Pre-Production Validation
- Post Insert Auditing

  NOTE: All validation results are recorded in the validation log DB tables.



Stage Validation

Pre Stage Validation

NACOR's incoming data currently comes from over 100 different sources. Data validation at this level is source specific. The validation process here is concerned about the structure and format of the data. The meaning of the data is taken into consideration in the second phase of validation (post insertion auditing).

1. **Data Type validation**
   Regardless of what is sent to NACOR all members of the same field must share the same "type."
   Whatever the data type (integer, decimal, and free-text) it must be uniform for all entries of the same field.

   Validation of all dates/times fields are correctly formatted as a date/time format.
   Validation of all numeric types (age, dose, concentration, etc.)

2. **Validation of ranges/formats**
   After a type has been determined the proper formatting and range checking can be applied. Some conditional logic may be applied at this time. For example:

   All ages must be integers (data type validation) and within a range of 0 – 110.
   All zip codes must be either five digits in length or five digits plus four separated by a hyphen.
   All anesthesia codes are character values and must be between "00100" and "01999".
   ASA Physical Status values must be between 1 and 6.

3. **Logic validation**
   With all fields verified to be of correct types and ranges verified some "business" logic regarding the data can be performed. This is also the point at which records missing crucial data elements may be deleted from the dataset.

   Admission date must be before the surgery date or on the same day.
   Start Time must be before the End Time
   The date of service (DOS) must be after the admission date.
   The discharge date must be on the same day or after the DOS. Checks for critical data: Anesthesia Type, Start / End Time, Providers, Facilities…

4. **Coding/Mapping Validation (referential integrity – Wikipedia)**
   The translation of source specific data elements to a single common format is the next step in the validation process. Values of data elements that are not currently present in our common definition libraries are flagged and scrutinized for accuracy. If the new value is valid it is added to our common definitions and assigned a proper code. Values that are determined to be invalid are removed and if the data element containing it is determined to be crucial, the record in question will be rejected.
   Coding/Mapping validation safeguards and referential integrity of data in NACOR serves as discovery process of new values. For example, if a data source which provides medication contains a new drug that is not currently part of the collected medications NACOR is currently aware of, a subject matter expert would then determine if data about the medication is valid and update the database with the appropriate medication definitions and codes for the new drug.

Sample of the mapping fields include:
Anesthesia type
US States
Gender
Measurement Units
Admission Type
Staff Role
Coverage Type
Timing Events
Outcomes
Drugs and Drug mixtures

5. **NACOR Specific Validation**
   Before data can be placed into NACOR certain pieces of information must be known. All providers and facilities referenced by a case must exist in NACOR. Before data is collected, an AQI participant is required to fill out a survey of their providers and facilities which is used as a reference point for case data.

Post insert auditing

All records in the registry undergo continual auditing on a daily basis. These audits include redundant checks from the pre-insertion validation routines as well as validations based on historically collected data. This trend analysis is performed in two different ways. The first involves comparing data over case time (the time at which anesthesia was performed for each case). The other involves comparing case data by the date it was imported into NACOR.

Comparing case data over time helps to determine the consistency of data from any particular data source. Key data elements in NACOR (# of cases, emergency status of cases, etc.) are cross checked with case time and rates are generated by month. Any unexpected conflicts yielding high variability are flagged and scrutinized by AQI staff.

Comparing data by insertion date helps determine the reliability of any particular data source. Again, key data elements (ASA physical status, patient age, etc.) are crossed with the date they were inserted into NACOR. An inventory of the range of enumerated values by data element and insertion date is compared across all insertion dates. A data source that has discrepancies by insertion date may need to be looked at more closely (such as ASA Physical Status).