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 the "messy" or bad data that could be introduced. 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 staging and validation (at load, for load)

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.

CPT or patient demographic information such as age is a good example. Is the 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? All of those automated tests must pass satisfactorily in the staging process before data is permitted to become part of NACOR.

Post insertion auditing (daily, all NACOR)

NACOR runs a daily automated (scheduled at night) 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) , like, say, Anesthesia Type, Gender, Timing Event.

Validation complete

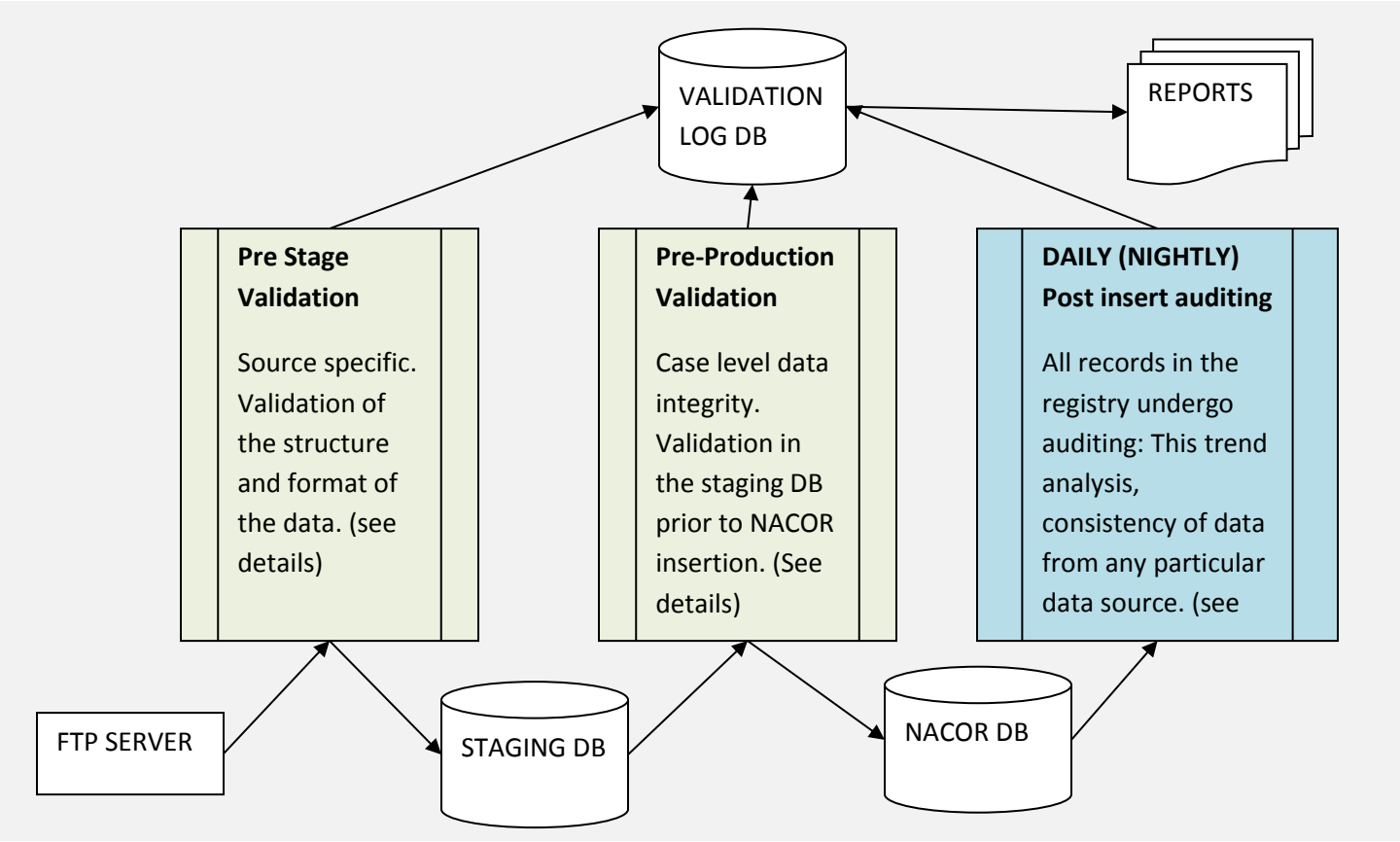
All case data is subject to intense scrutiny upon entry into NACOR. This is supplemented with a continual, more focused authentication process that last the entire lifetime of the data within the registry. With a two tiered validation method NACOR retains its accuracy while maintaining its ability to 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).
All steps of the pre-stage validation are recorded in a log table.

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, free-text) it must be uniform for all entries of the same field.

   Validate if all dates/times fields are correct are correctly formatted as 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.

   All ages must be integers (data type validation) and within a range of 0 – 110.
   Zip code.
   All anesthesia codes are character values and must be between "00100" and "01999".
   ASA Physical Status

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 <= the surgery date.
   Starts Time < End Time
   DOS > Admit Date
   Discharge Date >= DOS
   Records missing required data elements: Anesthesia Type, Start/End Time, Staff

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 bogus 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 referential integrity of data in NACOR and serves as discovery process of new values, eg.A data source providing medications contains a new drug that is not currently part of the collected medications NACOR is currently aware of. The subject matter expert would then determine the medication to be valid. The common definitions are then updated and an appropriate code is assigned to the new drug.).

Sample of the mapping fields include:
Anesthesia type
US States
Gender
Units
Admission Type
Staff Role
Coverage Type
TimeEvents
Outcomes

5. NACOR Specific Validation

Before any data can be placed into NACOR certain information must be known. All providers and facilities referenced by a case record must exist in NACOR.

All steps of the pre-stage processing are recorded in the log table for each load.
Sample of the one load process entries are in excel document

Stage pre-Production Validation

Pre-Production is an automated process (stored procedure with a series of tests) for case level data integrity check in staging prior to NACOR insertion.

The validation results are stored in log table.
Sample entries from the log DB

Stage	Test	Result	Errors
StageAIMS	Equal rows from table to view(PreOPLab)	PASS	0
StageAIMS	Equal rows from table to view(IntraOp IN/OUTput)	PASS	0
StageAIMS	Equal rows from table to view(IntraOpMed)	PASS	0
StageAIMS	Equal rows from table to view(IntraOpMonitor)	PASS	0
GENERAL	Stage Cases and Demographics have an equal amount of rows	PASS	0
GENERAL	NUM OF CASES WITHIN 20% OF NORM	PASS	8151
StageProcedure	ProcedureStatusFK - (within range 1 - 3)	PASS	0
StageProcedure	All records have parent in StageAnethesiaCase	PASS	0
StagePreOP	All records have parent in StageAnethesiaCase	PASS	0
StagePaymentInfo	All records have parent in StageAnethesiaCase	PASS	0
StageMap_Procedure_CPT	All CPT formatted correctly	PASS	0
StageMap_Procedure_CPT	All records have parent in StageProcedure	PASS	0
StageMap_PreOP_ICD	All records have parent in StagePreOP	PASS	0
StageDemographics	[AGE] - (within range 0 - 110)	PASS	0
StageDemographics	All records have parent in StageAnethesiaCase	PASS	0
StageAnesthesiaType	All cases have 2 or less types of Anesthesia Administered.	PASS	0
StageAnesthesiaType	All records have unique Anesthesia Types	PASS	0
StageAnesthesiaType	All records have parent in StageAnethesiaCase	PASS	0
StageAnesthesiaStaff	All AnesthesiaStaffID have parent in correct practice	PASS	0
StageAnesthesiaStaff	All records have parent in StageAnethesiaCase	PASS	0
StageAnethesiaCases	All AnesthesiaStartTime have value greater than 1/1/2010	PASS	0
StageAnethesiaCases	[CoverageTypeID] - (within range 1 - 7)	PASS	0
StageAnethesiaCases	[FacilityID] - (All Cases have a facilityID)	PASS	0
StageAnethesiaCases	[FacilityID] - (All Facilities is in DB)	PASS	0
StageAnethesiaCases	[AnesthesiaStartTime] - (duration of procedure < 24 hours)	PASS	0
StageAnethesiaCases	[AnesthesiaStartTime] - (start time < finish time)	PASS	0
StageAnethesiaCases	[AnesthesiaAQIID] - (ALL IDs are unique)	PASS	0
StageMap_StageProcedure_ASACPT	ALL ASACPT formatted correctly	PASS	0
StageMap_StageProcedure_ASACPT	All records have parent in StageAnethesiaCase	PASS	0

Post insert auditing (all NACOR cases as well as drill down to practice level)

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 case 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 loaded.

Comparing data over case 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 crossed with case time and rates are generated by month. Any crosses yielding high variability where it is unexpected are flagged and scrutinized by AQI staff.

Comparing data by load 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 loaded into NACOR. An inventory of the range of enumerated values by data element and load date is compared across all loading dates. A data source that has discrepancies by load date may need to be looked at more closely.

In either type of the daily post insertion audit trend analysis validation techniques, whether no flags are raised or potential problems are detected, a log is created to reflect the current status of NACOR's health. This log and reports is generated, read, and acted on upon daily.

Validation tests are added on continuous basis.

Current tests:

Validation Category	Validation SubCategory	Validation Description
AnesType	Orphan Anestesia	Anesthesia Type entries not associated with any Anesthesia Case
AnesType	Missing Anesthesia Type	Cases with no Anesthesia Type Specified
AnesType	Duplicate AnestType Records	Duplicate Anesthesia types in the AnesthesiaType table
Cases	Missing Months	Missing Months per Practice
Cases	Duplicate Loads	Number of cases per month is unusually high - indication of multiple loads.
Cases	AnesthesiasTime	Anesthesia Start Time is after Anesthesia End Time
Cases	AnesthesiaStartTime	NULL || < 2010 || > current
CPT	Cases w/o CPT	Cases in NACOR that do not have an entry in the map_procedure_cpt table
Demographics	Duplicate Demographics Records	Duplicate Demographics Records
Demographics	Orphan Demographics	Demographics entries not associated with any Anesthesia
Demographics	Age Missing	Age is missing from demographics
Facilities	PracticeID Missmatch	Facility (Facility PK) is assigned to different practices between WEB and NACOR
Facilities	Orphan Facilites	Orphan facilities - Facilities in anesthesia record (FacilityID) points to non existing facility in PracticeFacility table.
Facilities	CaseFacilityPractice Missmatch	Based on FacilitiesID - Anesthesia case points to different Practice than Facilitie associated with the case.
Facilities	Case - NULL Facilities	Anesthesia Case has no facility associated facility.
PreOp	ASA Class Missing	Missing ASA classes (from I to IV) in the last load.
PreOp	Orphant PreOp	Orphan Pre-Op - PreOp anesthesia fk points to non existing anesthesia record.
Procedure	Multiple Procedures	Duplicate Procedures in the Procedure table
Staff	Dormant Staff	Staff with no cases assigned to them
Staff	Multiple Staff ID	Multiple staffID for a practice. Indicating multiple entries for the same provider in the PracticeStaff table.
Staff	Cases w/o Staff	Cases in NACOR that do not have an entry in the AnesthesiaStaff table meaning there are no providers assigned to the case
Staff	Staff Orphan in Nacor	Staff in NACOR with no entry on the WEB practicestaff table.