

Data Quality and Consistency

The ultimate reward for participating in the STS National Adult Cardiac Surgery Database is meaningful data output to improve patient care delivery. This is achieved by building on good, complete data abstraction using STS data variable definitions, accurate data entry, data validation and data cleaning by querying and cross-checking your data. This section discusses some resources you can use to query and cross-check your data.

Querying Your Database

The end result of data entry into the STS database is data output and reporting at the database participant level and submission of data to the Duke Clinical Research Institute (DCRI) via the harvest process. Thanks to the hard work of the STS certified software vendors, the harvest process for all, or most, of the STS certified software products is automated.

Database participants engage in a varying amount of local reporting and output for both care improvement and data quality improvement purposes. Some participants choose not to do any reporting at the site level. Others get data out of their STS software as often as every 2 to 4 weeks. However your organization chooses to report data, to get useful, meaningful data out of your software, some querying of the data must be performed. To query your STS software means to set up criteria, filters, or conditions to answer a question. While some STS data participants might have data managers that are adept at pulling together queries using the local data collection system, others may need to work closely with their STS software vendor to get access to the information they need.

Technical aspects aside, to best ask questions of your STS data you will need to be familiar with two fundamental documents about the STS Adult Cardiac Surgery Database version 2.61:

- Data Collection Forms (annotated and non-annotated version)
- Data Specifications

These documents are included in this Training Manual.

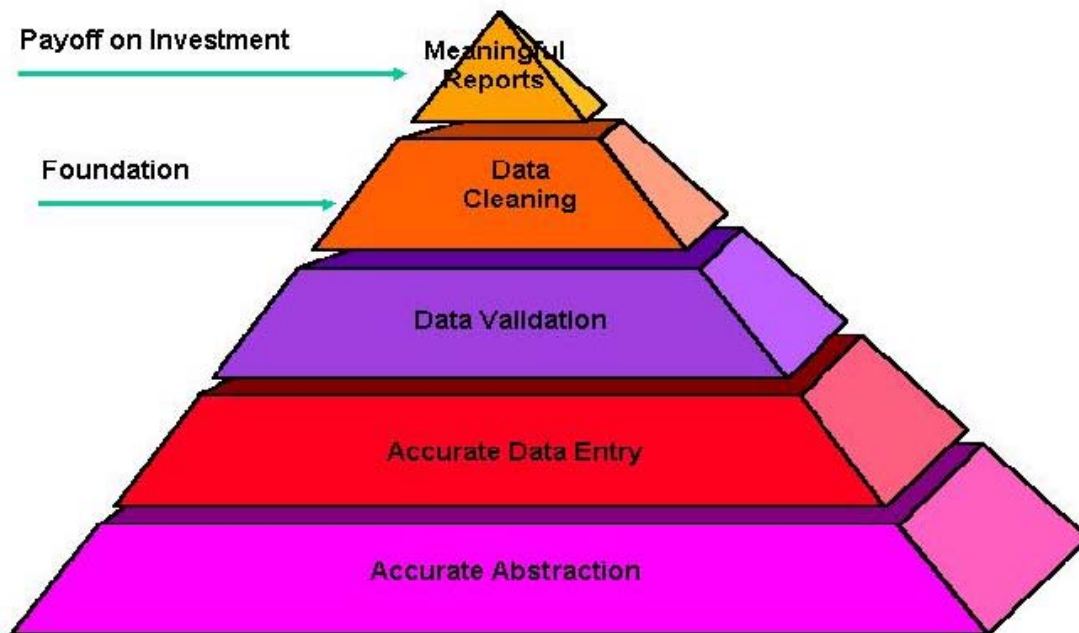
Another important tool that should drive your local querying efforts is the Data Quality Report (DQR) you receive back from the DCRI after submitting a data file during data harvest. This document contains detailed information about potential data quality problems that were identified at the DCRI data warehouse. Local queries of your data using the information in the DQR could help you quickly resolve any potential issues.

Although the technical aspects of querying varies across STS certified vendor software packages, data managers at other institutions could be another valuable resource to help you focus ideas for local data exploration efforts. The annual STS Advances in Quality and Outcomes meeting attended by the majority of STS database participant organizations is a venue where such collaboration and idea-sharing is common.

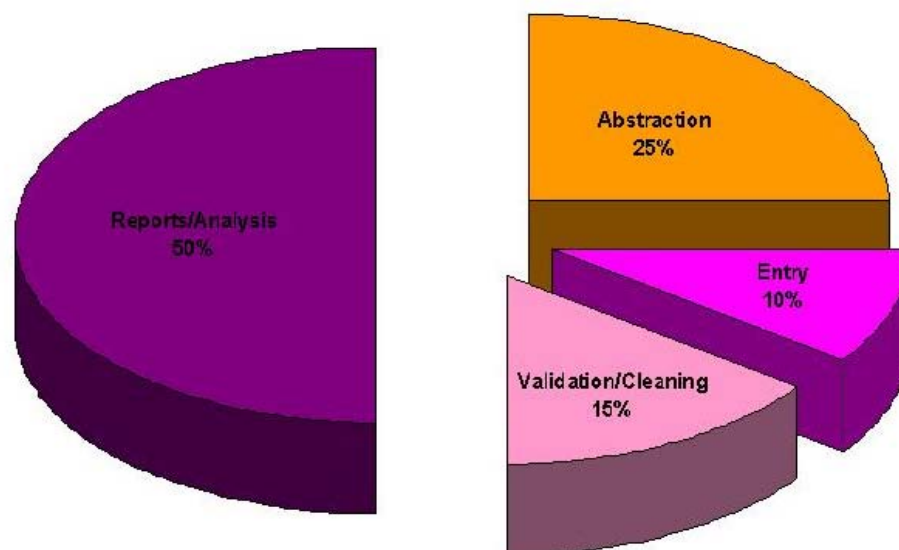
Data Elements to Cross-check

The ultimate reward in participating in the STS Adult Cardiac Surgery National Database is meaningful data output in some form of a report to improve patient care delivery. To obtain the ultimate reward, you must have accurate abstraction using STS data element definitions, accurate data entry, and data validation and data cleaning by cross-checking your data as your foundation for meaningful data output.

Data Quality and Consistency



Time Management for the STS database at the site level:



Although Validation/Cleaning should represent 15% of the total time allotted for STS database management, it is a vital step to produce accurate reports/analysis, where half of the total time allotted for STS database management is represented.

Data Quality and Consistency

Categories for Data Cross-check

There are at least four (4) categories to consider for data cross-check and for each there is a corresponding document that should be referenced. The documents described below are also included in this Training Manual.

(1) Case volume by operative category (STS Procedure Identification Table v2.61)

The STS Procedure Identification Table lists all the criteria needed to query your STS data by the eight (8) major operative categories – CAB Only, AV Replace, AV Replace+CAB, MV Replace, MV Replace+CAB, AV Replace+MV Replace, MV Repair, MV Repair+CAB. Whenever data are pulled by operative category at the database participant level, these criteria should be used in order to compare to the STS regional and national benchmarks.

(2) Parent-child relationships (Parent Child Fields v2.61)

The Parent Child Fields document lists all of the parent fields with their corresponding child fields. The document presents the fields in the same order as the STS Data Collection Form and also contains variable short names and sequence numbers for convenience.

(3) DCRI consistency edits (Consistency Edits v2.61)

Despite the safeguards built into the STS certified software packages and the data cleanup efforts by database participants leading up to and during the data harvest period, there are inevitably inconsistent values in the data submitted to the DCRI that must be resolved using consistency checks and edits. These checks and edits are performed by the DCRI as part of the harvest process. All data inconsistencies that will require an edit are reported back to the site in the harvest DQR. Participants have the option of resolving data inconsistencies and resubmitting their data harvest file.

Every participant's goal should be to minimize the number of consistency edits performed on their submitted data. Data changed as a result of the consistency edits may affect the reported results as well as the STS risk modeling. As a result, results in your report will not necessarily match similar results that you calculate locally.

In an effort to reduce the number of needed data consistency edits, we recommend that you pay special attention to the data variables involved in the consistency edits throughout your data collection and data cleanup process. A dated document entitled Consistency Edits v2.61 will be available as part of this Training Manual as we get closer to the first harvest containing data version 2.61 records. **Please note that this document will be updated through time if any new inconsistency checks and edits are identified. Please make sure you are referencing the most up-to-date version of this document.**

(4) Variables involved in STS risk modeling (STS Risk Model Variables v2.61)

The STS Risk Model Variables table indicates all of the STS v2.61 variables that are involved in STS risk modeling. This list should allow you to focus and prioritize your data cross-check efforts if time dedicated to the STS database is an issue.