

Will my SDTM and ADaM Support my Submission?

Nate Freimark
e-mail: Nate.Freimark@TheoremClinical.com

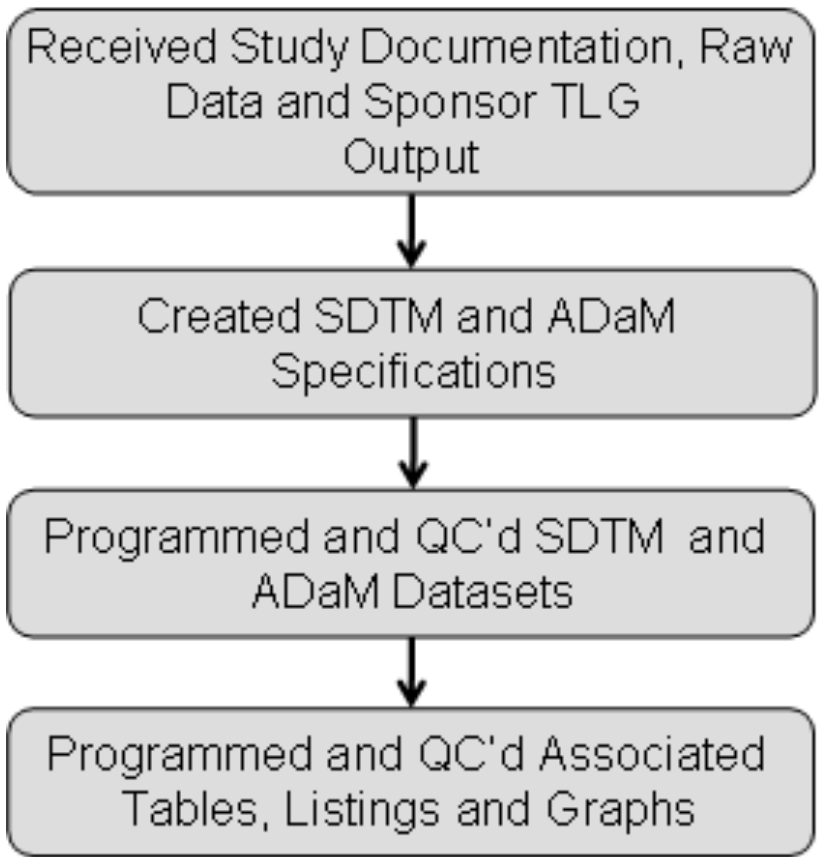
Victor Falch
e-mail: Victor.Falch@TheoremClinical.com

Theorem Clinical Research
1016 West Ninth Avenue
King of Prussia, PA 19406

ABSTRACT

In the imperfect world where the data flow is not SDTM to ADaM to CSR, the production of a CSR is disconnected from the process of creating CDISC compliant data. Companies often end up creating SDTM and ADaM right before submission to comply with reviewer expectations. The issue that arises is that the outputs and results provided in support of the submission may not be reproducible from the submitted SDTM and ADaM. This poster will address the various considerations and pathways that may be followed to assess what steps (if any) should be taken in order to improve the quality of the submission package. Several scenarios will be described along with the actions taken as a consequence of the findings that may arise. The various checkpoints (dataset checks, output re-QC) and methodologies that have been used will be described. The following represents experiences encountered while attempting to recreate non-CDISC compliant studies.

BRIEF OVERVIEW OF TLG RECREATION PROCESS



At the start of the study, important documentation such as the SAP, Protocol, CRF, CSR and sponsor output were obtained electronically, as well as the Raw Data. To supplement the study documentation, the sponsor had also sent SAS programs for the previously created datasets and output. After receiving the necessary study materials, SDTM and ADaM specifications were created according to the associated Implementation Guides. In this process, the CRF, Protocol and SAP were referenced in order to ensure correct data mapping, and to assist in the derivation of variables. Once the specifications were completed, both SDTM and ADaM datasets were produced. During the programming process, sponsor datasets were referenced, in order to preserve consistency in the representation of the data. After the SDTM and ADaM datasets were completed, TLG programming began. In this stage, sponsor outputs were used as mocks, and all TLGs were programmed to match the sponsor's.

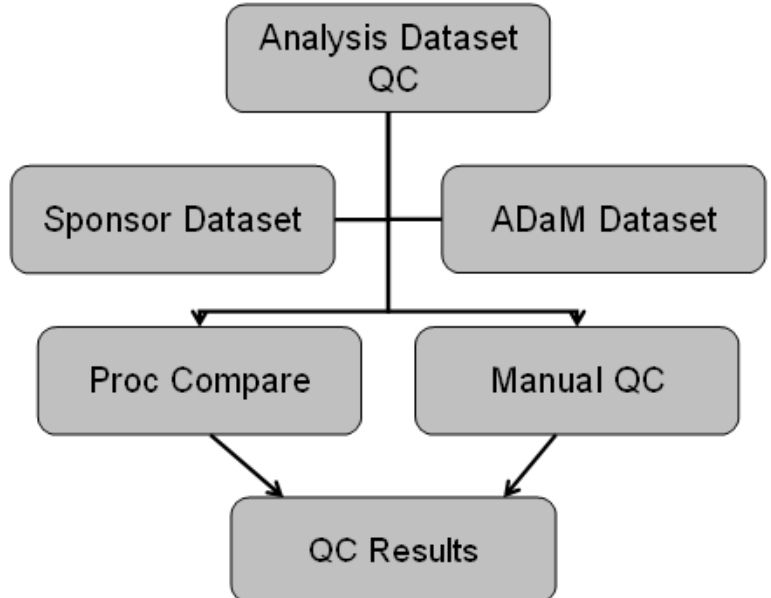
QC METHODS USED TO COMPARE SPONSOR OUTPUT

QC Method:	Analysis Datasets	Tables	Listings	Graphs
Proc Compare	✓	✓		
Automated QC		✓	✓	
Manual QC	✓	✓	✓	✓
Associated Table Comparison				✓

Analysis Datasets

➤ **Proc Compare**
Proc Compare was used to compare the sponsor datasets to the datasets we produced. The structure of the datasets was compared, as well as the number of observations. All data points were compared to ensure that derived variables had matching values.

➤ **Manual QC**
If there were any discrepancies during the dataset comparison, then the issue was investigated and Manual QC was issued.

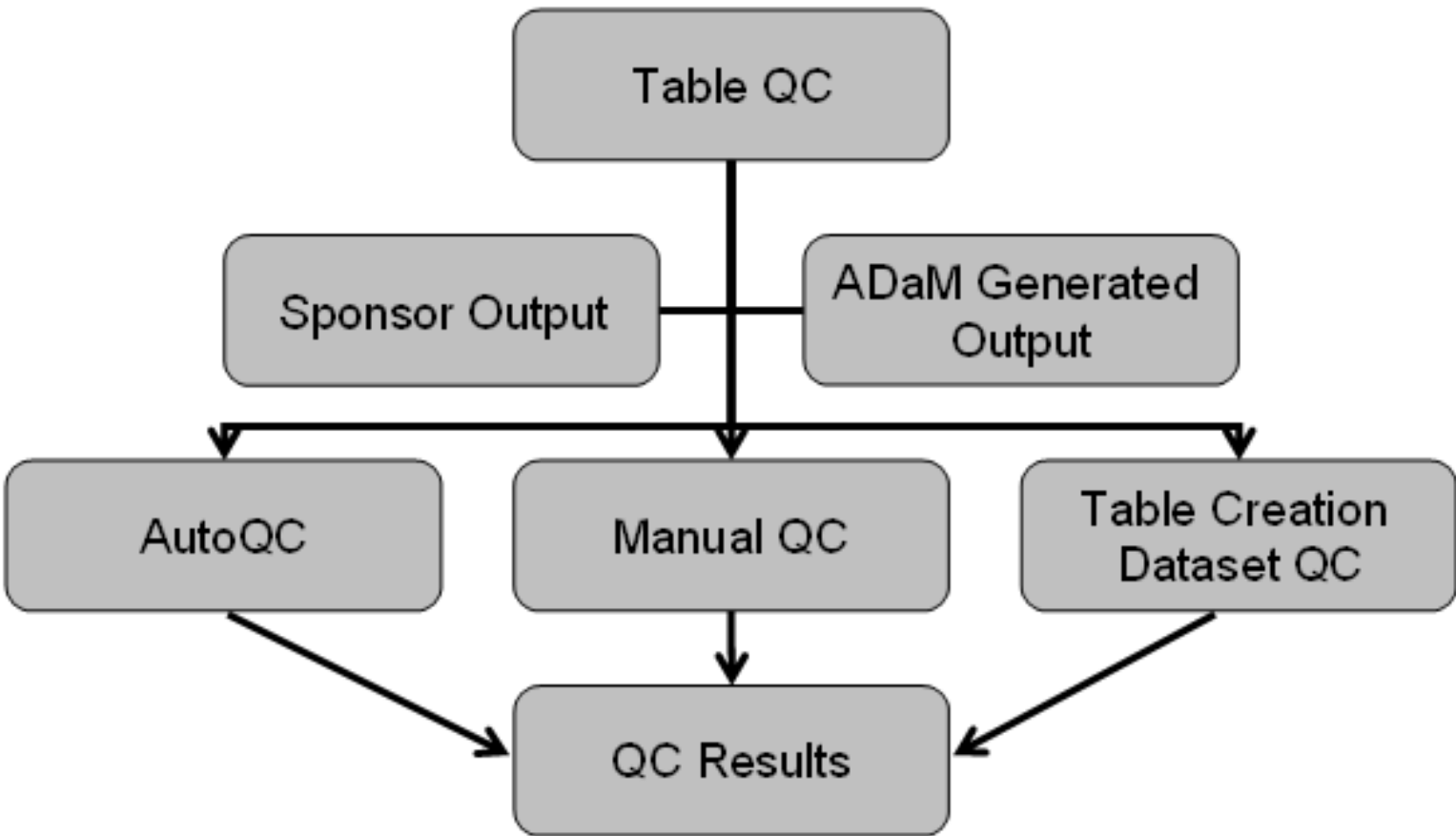


Tables

➤ **Automated QC**
AutoQC was implemented to compare output generated from ADaM, to sponsor output. This is a cell by cell comparison that is executed for each table. Each number in the table must match up 100% for there to be no discrepancies.

➤ **Manual QC**
If the numbers did not match exactly, then the issue was investigated and Manual QC was issued.

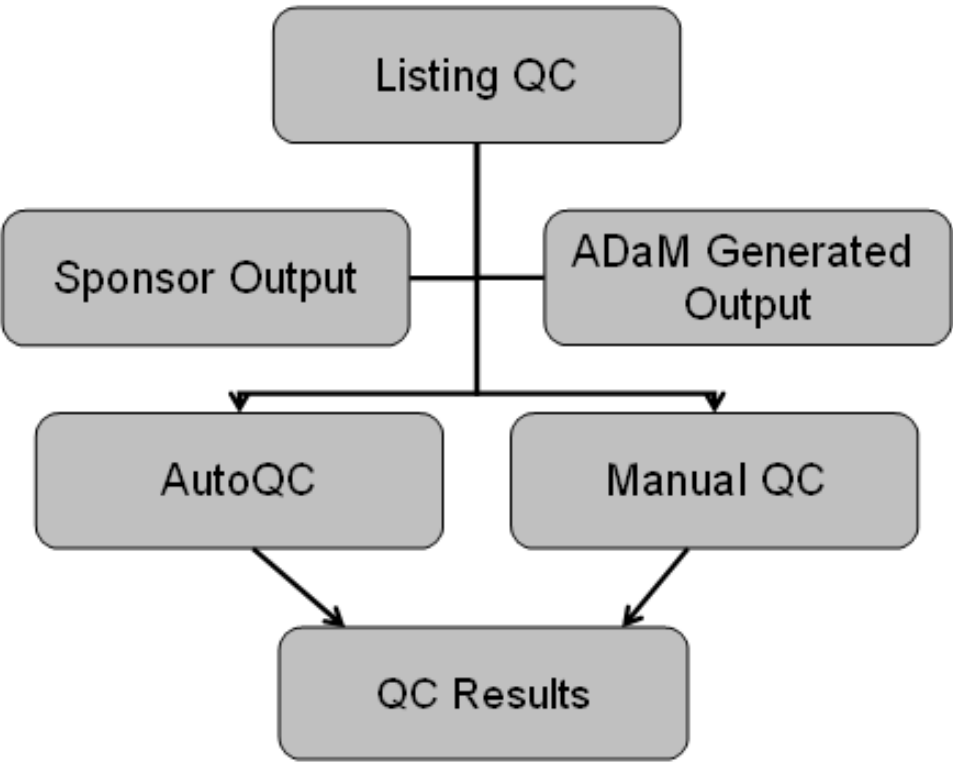
➤ **Table Creation Dataset QC**
As an alternative to table output qc, the table creation dataset qc was also used. This process compared the datasets used to produce each table. A cell by cell comparison was performed to ensure that all data values matched exactly. Table Creation Dataset QC is another form of automated qc.



Listings

➤ **Automated QC**
AutoQC was implemented to compare output generated from ADaM to sponsor output. This is a line by line comparison that is executed for each listing. All text in the listing must match up 100% in order for there to be no discrepancies.

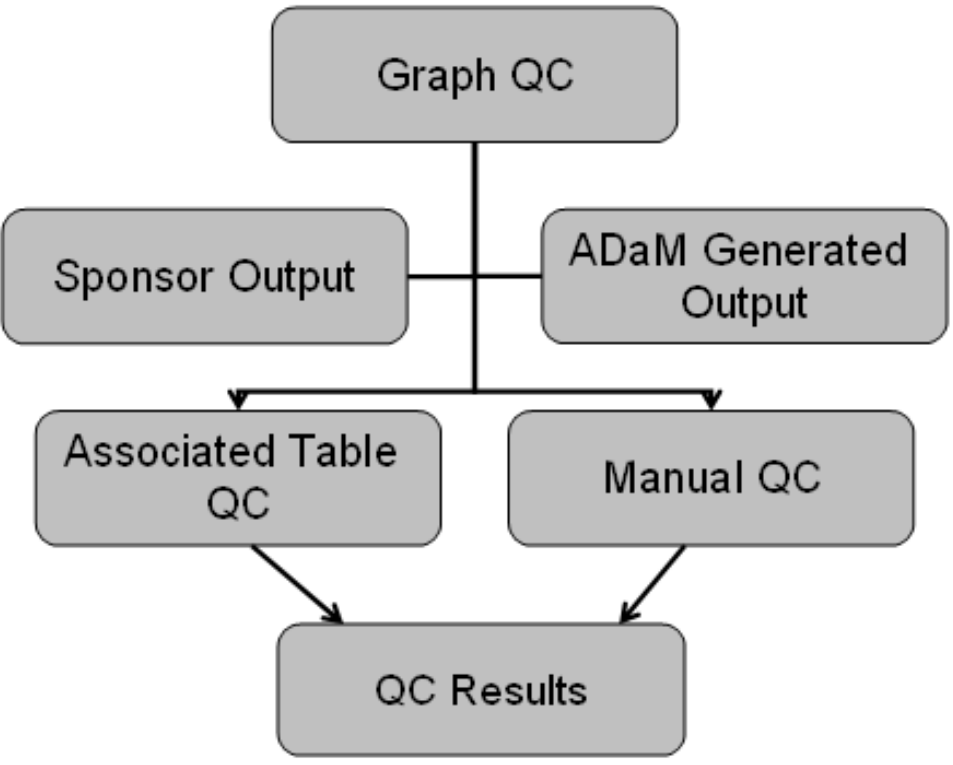
➤ **Manual QC**
If the text did not match exactly, then the issue was investigated and Manual QC was issued.



Graphs

➤ **Associated Table Comparison**
To make sure the output matched the sponsor's, we used the sponsor's associated table to qc our version of the graph. Each data point on the table needed to match what was being presented on the graph.

➤ **Manual QC**
If the data points did not match exactly, then the issue was investigated and Manual QC was issued.



Communication of Issues to Sponsor

Throughout the study, any issues noted had been communicated with the sponsor by use of a discrepancy log. The discrepancy log was an excel document that contained any issues discovered while working on the study. This file was sent to the sponsor for review, and either an updated file or email was sent back for any required updates.

Output Name	Issue Description	Reported by	Date Identified	Resolution	Status
Table 1 Disposition	Sponsor has value '0' for all groups in row 'Completed Study' and 'Primary Reason for Study Withdrawal'. We usually do not display values if they are '0'.	John Smith	21-Aug-2011	Do not display the values if they are '0'.	Resolved
Table 5 Electrocardiogram	Values do not match on baseline visits. Sponsor rounded original collected data, then calculated average, then rounded the average, then reported baseline values. We rounded after all calculations were completed.	John Smith	21-Aug-2011	Use the rounded values after all calculations are completed.	Resolved
Table 6 Chemistry Labs	Sponsor has double parenthesis in the units for Creatinine Clearance '(mL/min)'. This should only have one pair of parenthesis.	John Smith	23-Aug-2011	Only use one pair of parenthesis in the table.	Pending
Listing 10 Randomization	Sponsor failed to subset data to subjects that have randomization flag set to 'Y'.	Jane Doe	21-Aug-2011	Subset data to subjects with randomization flag equal to 'Y'.	Resolved
Listing 15 Lab Reference Ranges	Sponsor is displaying 'TSH' under 'Chemistry Tests'. This should be displayed in the 'Other Tests' section of the listing.	John Smith	24-Aug-2011		New
Listing 15 Lab Reference Ranges	Hematocrit and Hemoglobin should be under 'Hematology Tests'. Sponsor has both under 'Chemistry Tests'.	Jane Doe	24-Aug-2011		New

ISSUES DISCOVERED IN THE RECREATION PROCESS

Handling Baseline and Unscheduled Visits

Throughout the study, there were issues regarding whether or not to include unscheduled visits in baseline calculations. The SAP was not clear with how baseline was to be calculated. As a consequence, this issue caused discrepancies between our output and the sponsor's. As a result, we had to work very closely with the sponsor to agree on baseline calculations.

Inconsistent File Formats Received from Sponsor

When the study documentation was received from the client, it was discovered that we were provided different file formats for some of the outputs. This made the QC process for the outputs difficult because the AutoQC method was not supported by certain file formats. As a consequence, a lot of time was invested in working with the sponsor to convert the files to a supported format. As a result, if we were unable to obtain new file formats from the sponsor, we had to modify our QC programs or find a way to convert the files ourselves, in order to work around the issue.

Different File Formats Received from Sponsor

- **Microsoft Word**
- **Microsoft Excel**
- **Adobe PDF**
- **Text**
- **Delimited (Comma, Tab, Etc.)**

Rounding Issue in the Derivation of Variables

It can be difficult to recreate data because new variables may rely on other variables in its derivation. When creating a new variable, rounding may have occurred throughout the process and it may be very difficult or impossible to recreate this exactly. As a consequence, mismatches will occur between the output. As a result, variable algorithms need to be detailed and confirmed by the sponsor.

Sponsor Dataset:

PATNO	VISIT	TEST	RESULT	UNIT
001	Screening	BMI	25.7	KG/M2
009	Screening	HR	60	BEATS/MIN
152	Screening	BMI	22.3	KG/M2
383	Screening	HEIGHT	180.2	CM

Recreated SDTM Dataset:

USUBJID	VISIT	VSTESTCD	VSORRES	VSORRESU
001	Screening	BMI	25.8	kg/m2
009	Screening	HR	60	BEATS/MIN
152	Screening	BMI	22.5	kg/m2
383	Screening	HEIGHT	180.2	cm

Sort Issues in Listings

Since we were trying to match output exactly, we needed to make sure that not only our data matched the sponsor output, but the sort of the data did too. The problem with this, is that sometimes there is not a specific sort that is used. As a consequence, it is the luck of the draw how the data appears in the dataset. As a result, some listings needed to contain a hardcoded sort in order to match sponsor output.

CONCLUSION

Overall, the process of recreating non-CDISC compliant studies was difficult and many obstacles were encountered along the way. New procedures and methodologies had to be created in order to successfully complete the data conversion. We hope that this poster has provided useful information and would ease this process for future work. If you have any questions, please feel free to contact Nate Freimark or Victor Falch at the email addresses stated above.