

QC—Manual vs Program

Helen Nicholson, Cmed Clinical Services



Introduction

When an output (dataset, table, figure or listing) is produced you need to know that it is correct– you need to quality check (QC) it.

There are lots of different ways to QC, and the method you use will depend on what you are producing. You need to check the output you are producing is what is expected. For example there is no good producing a demographics table when it should have been an adverse event table. You also need to check that the contents of what has been produced are correct.

One of the biggest questions is how do you do the checking – 'manually' or electronically? The reality is usually a bit of both. The questions how much is done programmatically or electronically, and what and how it is done can take a lot of time and effort to answer.

Your company/organisation will have standard operating procedures (SOP's) that specify what is needed to prove that you have done the QC. They may specify the QC method or they may leave it up to you.

The following assumes that the original programmer has already checked the output and it is an independent person doing the QC

Table, Listings and Figures - Manual checking

There are always some manual checks to do –

- Does the output match the shell?
- Does it look "right"?
- Are the footnotes/titles correct?
- Are there any missing values which should be present?
- Are all the formats right, or are there some that are too small?



One check that is normally done manually, is consistency across tables – checking you have the same number of subjects in each table for example.

You can check the contents of the output manually, but it may take a long time, and it needs to be done every time the output is produced. Also, how do you check any statistics – do you sit with a listing entering the values into a calculator and hope you have found the same number the table has?

When you have checked something how do you prove you have checked it? The usual thing is to print it out, and tick what has been checked, and sign the output to say it has been checked. As this has to be done every time the output is produced for checking, you can end up a lot of printouts that need to be stored.

Table Figures and Listings - Electronic checking

For output that may have to be checked several times some form of electronic checking may well save time depending on how it is done.

There are three main types of electronic checking

- Dual program the data but not the layout. This gives you the same information but not in the same format layout which means that you have to then manually check the QC output against the original.
 - These programs can be very quick to write
 - The actual checking against the original can take time if the outputs are large.
- Dual program the data and the layout. You can completely dual code the output including the layout and then use some file comparison software to compare the QC output with the original.
 - The shells need to be very detailed to include the width of all the columns and where any white space is.
 - Most outputs have the run date/time on them, so you need to be able to ignore differences caused by this, or you would have accept then and sign them off as acceptable differences.
 - Normally there is a way of storing the comparison file electronically
 - If you are signing them off for large listings you may still have a large difference file which needs to be stored
 - What you save, or print and sign will depend on what is required by your SOP's.

Table Figures and Listings - Electronic checking (continued)

- Dual program the data with a programmed compare of the output. If you change the production program so that there is also a report dataset produced (out= option on PROC REPORT for example) and then dual code that dataset to get a complete match.
 - Each column across the page becomes a variable
 - Include the nonprinting columns
 - Set the labels to be the column headers
 - Each row down the page is an observation
 - The variable format will be the same as any given in the define statement
 - Standard naming conventions for the columns, including any ordering only columns makes things easier
 - Either drop the _break_ variables from the production dataset, or program them into the QC dataset.
 - For figures you may have to output the data that goes into the graph

Again you have the problem of proving what you have checked – for the first option it is normally print, tick and sign as for manual checking. For the other two if you keep the results of the electronic comparison and use some form of tracking which you sign at the end. If you have differences between the production and QC then you may have to print out the output and annotate on the reason before signing and dating. It all depends on your company SOP's.

Tables, Figures and Listings - Manual vs electronic

Some manual checking is always required, but to completely check the contents manually can be very time consuming every time the output is produced, although it may be the quickest method for a one off output that only contains a small amount of data.

Dual programming the contents only of the output can be quick for small outputs, but you need to manually check them against the original every time, which can be time consuming for large outputs.

Reprogramming the output completely and comparing with the original using some sort of the file comparison software can be time consuming if you have to get all the columns to match exactly. Then you need to decide what you do about the run date/times – ignore them, or accept them as differences?

Changing the production program to produce an output dataset and then dual programming that dataset needs more upfront planning. You need to know what to call the variables, and can take time to get columns to match, but once programmed it is very quick re run if you need to re run everything. By putting the return code of the PROC COMPARE into a dataset you can easily see the status of all the QC in one place

How you record what has been checked, and when will, depend on the company procedures.

Data sets - Manual checking

There are some manual checks that should still be carried out on data set

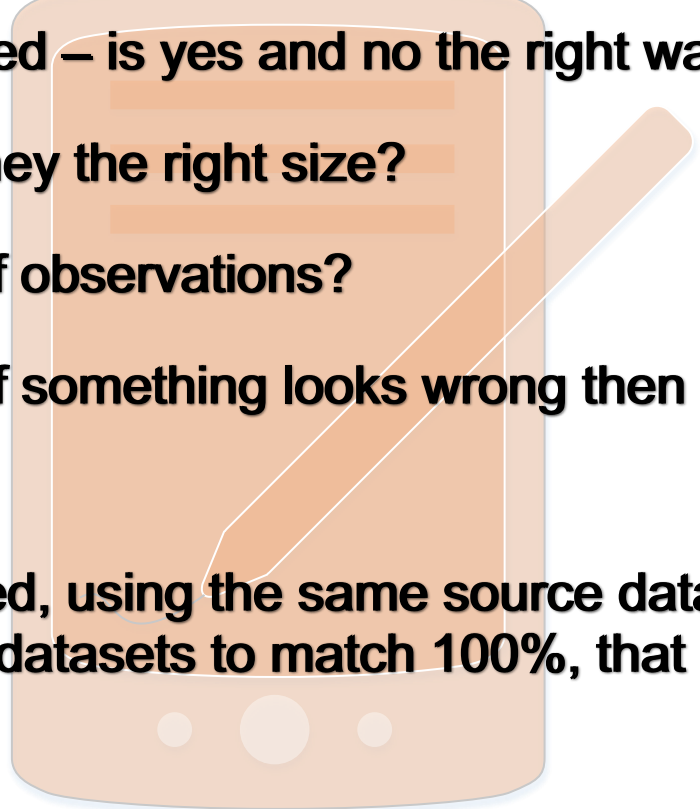
- Does the data fit the dataset – if a vitals signs dataset we don't want laboratory data in there,
- Have all the variables that need it got a decode for every value?
- Are the right decodes being used – is yes and no the right way round for example,
- Are the formats right and are they the right size?
- Are there the correct number of observations?
- Does the dataset 'look right' – if something looks wrong then it probably is

Electronic checking

Most dataset are dual programmed, using the same source data as the original production datasets. The idea is to get the datasets to match 100%, that is they match on

- The number of variables
- The number of observations
- The labels, lengths, informats and formats for all variables
- The content of the variables match

Although depending on your company procedures you may be allowed differences in some of the attributes.



Data sets - Manual vs Electronic

When checking the content of a dataset it much easier to do it electronically by dual programming of it. This means that when you get new data it can be quickly and easily re checked or re QC'd This is especially the case if the programmer has allowed for all possible cases ,and not just for the data that they currently have.

One advantage of aiming for a 100% match on everything is that if for some reason you don't match and it is know that the QC code is right, and there is something wrong in the production code, then you can swap the two over easily, so that the QC becomes the production code and the production the QC – you may then be able to sign off the differences.

However the dataset should still be looked at by eye to make sure that

- It looks reasonable,
- There are not a lot of duplicate observations that should not be there,
- If a variable or observation contains a lot of missing data, that it is reasonable, and not a result of the programming..

If the return code from the PROC COMPARE is stored in a dataset, along with the dataset name, then it is possible to see the complete status of the QC just by opening one dataset.

Conclusion - a personal point of view

When it comes to TFL's I have experienced a great many variations in what is checked and how it is checked. There always needs to be some form of manual checking – and there should always be that last manual sanity check before it goes out. For small one-off outputs I think the quickest way is to dual program the content but not the layout and then manually check the QC against the production.

For anything else my preferred way is, currently, to produce an output dataset and dual code that. This is made a lot easier if you have standards for the variable names, for example column1 for the first column, column2 for the second etc, and the variables are compressed, or left aligned, so that you do not have to match on white space.

For datasets, I think that they have to be dual programmed, and then compared. My preference is to get 100% match, although when I only differ on a few points in a dataset, and time is short, I would recommend manually checking that the production is right.

If you make sure the name of whoever is running the QC is recorded in the compare when it is done, together with the date and the result is stored electronically you have your proof. Adding information to a tracking sheet makes it easier to track, and gives a place to sign if required.

