Introduction

When an output dataset 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 we do the checking - manually or electronically? The reality is usually a bit of both. The questions that match data are done programmatically or electronically, and what information is done can take a lot of time and effort to answer.

Your company/organisation will have standard operating procedures (SOPs) that specify what is needed to prove that you have done the QC. They may specify the 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.

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

✓ Does the data file fit the dataset -- if a vital 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 datasets are dual programmed, using the same source data as the original. The idea is to get the datasets to match 100%, that is they match on the number of subjects in each table for example.

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 found the same number the table had?

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.

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 checking with the original using some sort of the file comparison software can be time consuming to match exactly. Then you need to decide what do you do about the run-time differences - ignore them, or accept them as differences?

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

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

Data sets - Manual checking

Table, Figures and Listings - 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 found the same number the table had?

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.

Manual checking

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

✓ Does the data fit the dataset -- if a vital 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 datasets are dual programmed, using the same source data as the original. The idea is to get the datasets to match 100%, that is they match on the number of subjects in each table for example.

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 found the same number the table had?

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.

Conclusion - a personal point of view

When it comes to TPL 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 enough 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 a lot easier if you have standards for the variables, for example column for the first column, column 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 productions right.

Table Figures and Listings - Electronic checking

If you make sure the name of whoever is running the QC is recorded in the compare when it is done, you can get columns to match, but once programmed it is very quick re run if you need to re run everything. By getting the return code of the PROC COMPARE into a dataset you can easily see the status of all the QC in one place.

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

Tables, Figures and Listings - Electronic checking (continued)

The shells need to be very detailed to include the width of all the columns and where any white space is.

Normally there is a way of storing the comparison file electronically

If you are signing them off large listings you may still have a large difference file which needs to be stored.

Table, 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 to match exactly. Then you need to decide what do you do about the run-time differences - ignore them, or accept them as differences?

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

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

TFL's

When it comes to TFL 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 enough 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.

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 found the same number the table had?

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.

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 to match exactly. Then you need to decide what do you do about the run-time differences - ignore them, or accept them as differences?

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

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

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?