PhUSE US Connect 2018

Paper TT09

Generating tables, listings and figures without any programs

Shafi Chowdhury, Shafi Consultancy Limited, London, U.K.
Md Zobair, Shafi Consultancy Bangladesh, Sylhet, Bangladesh

ABSTRACT

The programming and validation of Tables, Listing and Figures for Clinical Trial Reports is an essential part of the reporting process to ensure integrity and quality of the analysis, and although technology has moved on over the years, this part of the reporting process remains untouched. Why do we still have to write programs for tables, listings and figures? If we just write programs to analyse the data as defined in the SAP and store the results, we can then use an application to present the results in the various types of tables, listings and figures. This means that each analysis is only performed once and this one result is used in the different outputs as required. This not only leads to CONSISTENCY and QUALITY within and across trials, it also significantly reduces the programming resource and time required.

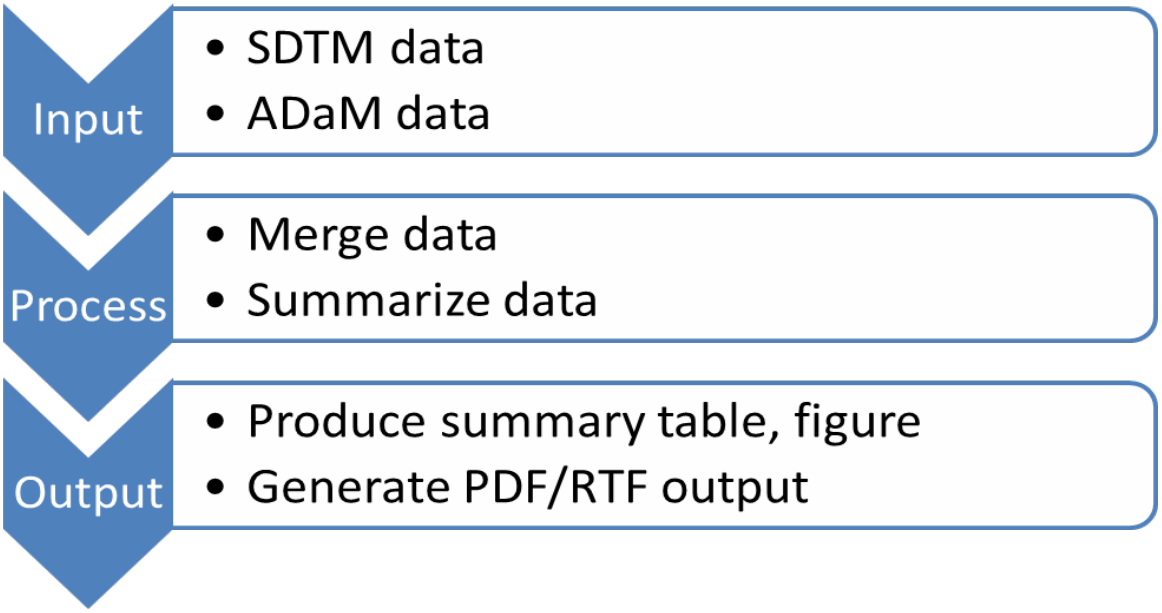
INTRODUCTION

Tables, listings and figures (TLF) are generated as part of the normal reporting process. The trial data is analyzed and presented as tables and figures in clinical trial reports. The tables, listings and figures are also used to answer regulatory questions and to support publications based on the trial data. One of the key aspects of any summary result is transparency. There must be a clear path from raw data to the summary result that is used in a table, listing or figure.

SAS programs have been the backbone of transparency over the years, and recently R and other languages are also being used to generate summary results used in trial reports, regulatory responses and publications. The transparency is captured by the use of a program, log and output, with data going into the program and the output being summary results that are displayed in a tabular format in an ASCII, PDF or RTF file. However, as technology develops and more tasks are controlled and managed by metadata, the requirement for a program to link the raw data to a summary table is no longer necessary. The same end result can be achieved by using a traceability matrix showing which raw data produced which summary result and in which table(s) that summary result is used.

PROCESS

In order to produce summary tables, listings and figures without any programs the process currently used must be analyzed and gaps identified. The output programs currently go through the following process:



METADATA

As technology evolves and metadata is being used more and more to control and manage the tasks required, from defining which table shell to use for each table, to which dataset and variables, it seems that we are increasingly duplicating our efforts. The metadata can be used to auto-generate programs which can then be used to generate the output. However, this is what has already been done a long time ago, and more should be done to solve the current problems.

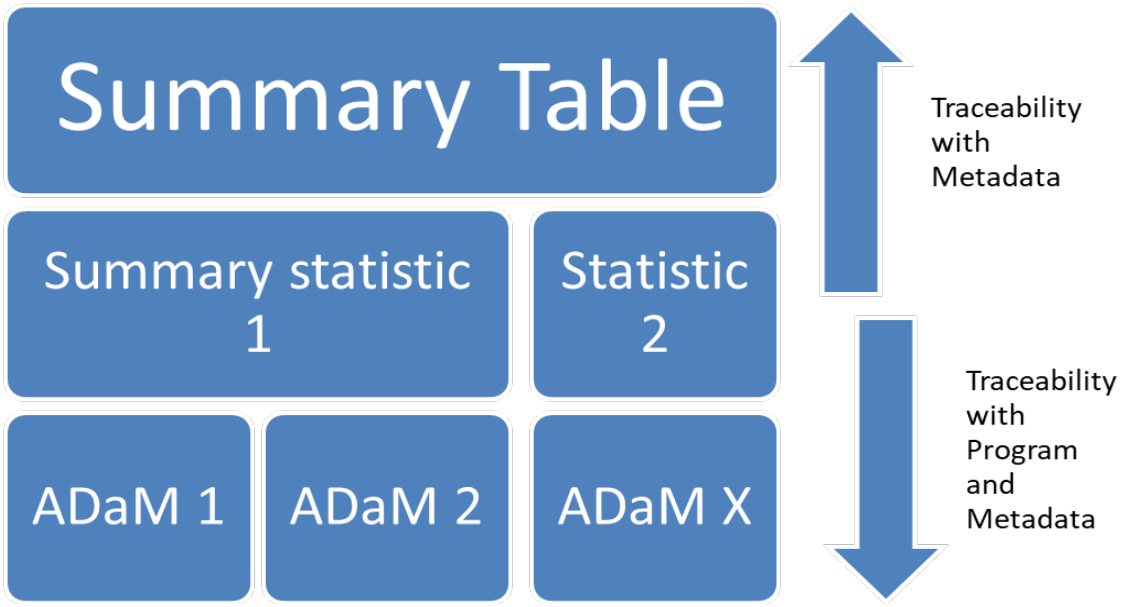
PROBLEMS

The problems in the current process are that in each trial there are some variations, and then the medical writers would like something different than what was produced for the trial report. A slight difference can mean extra time to rewrite the program and to validate it, or as in most cases, write and validate new programs.

Besides the time taken to produce extra tables, if there are minor changes in the output then the programs have to be updated, revalidated, and the output rechecked to ensure the numbers are consistent across tables. Something basic like number of patients in treated set can be calculated hundreds of times during the course of a trial. So it is vital that this is checked across all tables to ensure the numbers are correct and consistent.

BREAKDOWN

If the problems can be dissected and if we look at the current process, there are two major parts that needs to be addressed. There has to be traceability from rawdata to the result, and the result then needs to be used to propagate summary table or figure.



SOLUTION

If the tables and figures are split up so that the summary statistics are derived first as defined in the analysis plan and then the different statistics are put together to create a table as defined in a table shell, then a combination of current and new technology can be used to create a more efficient process, one that allows full traceability and full flexibility. By calculating the statistics first using a program based on metadata, then using more metadata to specify which result should go to which output table or figure, the user has full flexibility on the table shell without having to validate the numbers going into it each time.

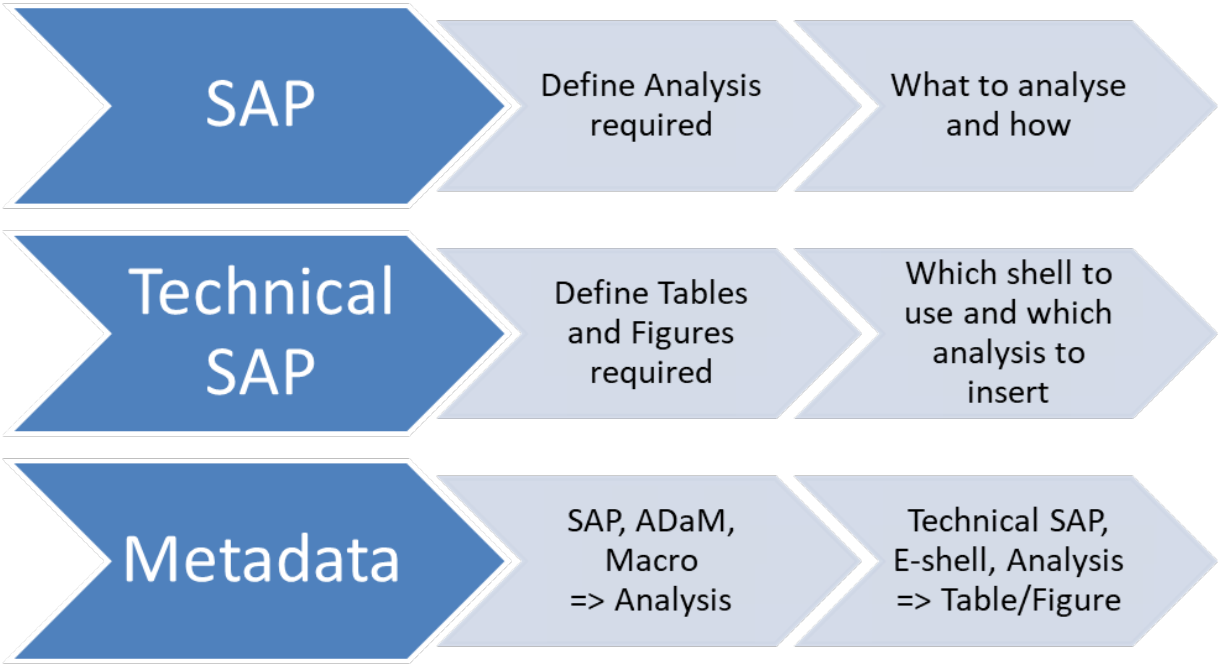
Although different organizations will generate the statistics separately, using metadata and automation technology means these can be standardized. This not only ensures that there is consistency in the programs used to generate the results, but there is consistency in the results used within different table and figure shells, thus minimizing the risk of inconsistency in numbers.

The generation of specific tables and figures using the result is the final part of the process. How can this be done without any programs? In the past the shells were mere ASCII files that then had to be used by a SAS program to be updated with the results. However, they can be electronic templates where an application uses meta data to insert

specific results into the shell. The application would be validated using software validation process to confirm that numbers being placed in the shells are not changing, meaning the process would be entirely metadata driven.

DATA FLOW AND CONTROL

Control of metadata is key for this process to work. It would use the information from the Statistical Analysis Plan (SAP) that defines which analyses are required to determine which ADaM datasets and variables are required together with which standard analysis macro should be used. This will then lead to programs that generate validated analysis. The second part of the metadata comes from the Technical SAP that defined which tables are required, which shells should be used and with which analysis. If all this is kept in a metadata database then full traceability is available from raw data to the final tables and figures produced.



FLEXIBILITY

Splitting up the process into deriving the analysis results and then producing the tables or figures means the layout can be changed without affecting the validated results, so no need to update programs. It also means that different groups can use the results, e.g. the Medical Writers can use the data to generate their own tables without there being any risk that the numbers used are incorrect. The results can also be presented in Figures or Tables without this needing to be pre-defined when the analysis program is generated.

PROGRAM STRUCTURE

As standard macros will be used by the metadata to auto-generate programs, it means the programs will follow good programming practice and have the same structure, from using a standard program header to the layout of macro calls. This will make them not only easy to follow by everyone, including anyone from the regulatory if the programs are submitted, but that they will be easy to update over time if required. Across studies the programs will be using the same company standard macros, ensuring that it is easy to compare which analysis was used in which study or with which type of endpoint. This will give a greater transparency than achieved in the past.

The programs generated from the metadata will have standard headers and each analysis will be commented and structured in an easy to read manner. An example of an analysis code could look like this below, but it can also take on any other structure as this is defined in the macro that auto-generates the programs from the metadata. Every analysis performed will be programmed in a similar manner regardless of whether a program contains one or more analysis.

```
**********************************************************************;
*** Output No.
*** Title: 3.2.9 Analysis of Age by treatment - TS
**********************************************************************;
```

```
FILENAME O '!baseloc\lst\anal_age_ts.lst ';

PROC PRINTTO PRINT=O NEW;
RUN;

%T_MSUM (  ANAL_TITLE    = 'Age [years]'
        , ANAL_NO       = 3.2.9
        , POPU          = TS
        , INDATA        = adam.adsl
        , CONDIN        = age LE 60 AND saffl EQ 'Y'
        , OUTDATA       = outcat_age
        , STATS         = N MEAN MIN MAX
        , TOTAL         = Y
        , SUM_TYPE      = nsum
        , STATVAR       = age
        , DECSD         = 1
        , SUM_TYPE_CODE = 2
        , BYNUM         = 0
        , COLVAR        = trtn
        , COLVARLAB     = trt
        , DECMINMAX     = 0
        );

PROC PRINTTO PRINT=PRINT;
RUN;

PROC PRINTTO LOG=LOG;
RUN;
```

TABLE SHELLS

The shells used for the tables and figures will be different from those used currently. As the final table will be determined by the analysis going into the shell, it means the shell will be more generic. The template 1 below can be used to produce many different tables from demography to adverse events and template 2 can be used for anything a summary table by sex for example to an efficacy table, simply by allowing data to drive what is going in each of the row or column variables.

Template 1:

Column Label (e.g Demographic Parameter)	Colvar1	Colvar2	Colvar3	...	ColvarN
Select Analysis Item					
Cat 1	XX.X	XX.X	XX.X	...	XX.X
Cat 2	XX.X	XX.X	XX.X	...	XX.X

Template 2:

Column Label (e.g Demographic Parameter)	Colvar1			Colvar2			...	ColvarN		
	col 1	col 2	col 3	col 1	col 2	col 3	...	col 1	col 2	col 3
Select Analysis Item										
By Var1										
By Var2	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	...	xx.x	xx.x	xx.x
By Var2	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	...	xx.x	xx.x	xx.x
By Var1										
By Var2	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	...	xx.x	xx.x	xx.x
By Var2	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	...	xx.x	xx.x	xx.x

RESULTS DATABASE

In order to effectively use electronic templates a standard results database is required. This should be structured so that all the different types of analysis, including those stratified by one or more covariates, can be stored without the need to alter the structure. This makes the results database a key component of this process, and time and care should be taken to define the best structure possible. If it can remain unaltered, then tables and figure templates can be generated without the need for updates when new analysis are performed.

CONCLUSION

With the advancement of technology from a greater use and understanding of metadata to automation of program generation, the days of programmers spending time writing programs for tables, listings and figures are numbered. Application generated tables, listings and figures are already accepted in many sectors, and although the pharmaceutical industry is coming late into this, provided traceability can be demonstrated from raw data to the analysis results that appear in any subsequent table, listing or figure, there is no reason why this process cannot be used to remove the need to program any analysis that has already been programmed once somewhere within an organization. Tables and figures can then be generated using these results and electronic shells.

Standardization is key to any process that can stand the test of time. In order to produce tables, listings and figures without any programs many parts of the reporting process need to be standardized. From the SAP it should be clear what data to analyze, the SDTM and ADaM structure will allow for analyses programs to be standard, the results from the analyses must be stored in a standard results database so that standard shells can be used to produce tables, listings and figures which are from a standard list of outputs.

Programmers have long since argued that there is not enough time to perform new exploratory analysis, however, if the programs are auto-generated based on metadata, then this will drastically reduce the time required to generate standard tables, leaving more time to not only review the data to ensure that it is clean but to also perform new and more innovative analyses, helping to deliver drugs quicker to the market.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Shafi Chowdhury
Shafi Consultancy Limited
7 Blossom Way, Uxbridge UB10 9LL
United Kingdom
Email: shafi@shaficonsultancy.com

Md Zobair
Shafi Consultancy Bangladesh
Sylhet, Bangladesh
Email: zobair@shaficonsultancy.com