

A Year in Pharma

Gemma Bishop, Quanticate, Manston, UK

ABSTRACT

Peruse SAS® programmer adverts on popular recruitment websites, and the adverts would read like this: "As a SAS Programmer you will provide SAS software programming support for all data management and analysis activities, and be involved with the maintenance and development of in-house SAS software tools to support database mapping, data management tasks and analysis tasks." The aim of this paper is to examine the traditional role of a SAS programmer and whether the balance of time spent manipulating versus analysing data is correct in the current Pharma climate. It will use the author's experiences working in other sectors to examine how the current situation was arrived at, and what benefit we gain from it. It will establish whether the current status quo to change – whether by staying as we are, we are losing out somewhere. And if change is required, what needs to happen to cause that change, and what the brave new world of SAS programming would look like afterwards.

INTRODUCTION

The role of programmer within the Pharmaceutical industry is now a well-established career in its own right. Programmers can now find themselves expected to perform a range of data management and analytical tasks. The establishment of PhUSE as a forum for programmers to develop, exchange ideas and experiences shows the level of commitment within the industry to ensure that programmers remain a vital part of the clinical reporting process.

This paper will examine the current role of a SAS programmer within the Pharmaceutical industry. It will look at how this role has developed, and whether, as the industry changes, how the role of the programmer needs to change to fit in with the new landscape.

PROGRAMMING IN THE PHARMACEUTICAL INDUSTRY - THE CURRENT SITUATION

A programmer within the current Pharma environment can expect their responsibilities to be around assisting in reporting clinical trials. The majority of the work involved is the production and validation of derived datasets, and the programming of tables, figures and listings (TFLs). However, my experience over the last year since joining the pharmaceutical industry is very different. Before a programmer is in a position to be able to produce TFLs, a not inconsiderable amount of time and effort is required to query, fix, and clean the data.

In addition to the production of said TFLs, a programmer may find themselves working on submissions to regulatory authorities. If this requires the pooling of analyses, conversion of data structure will be required to ensure that all the data follows a standard structure. Work on this type of study will inevitably require a higher proportion of time spent on cleaning, fixing and validating the data.

The diagram below is a very simple illustration of the role of a SAS programmer within the pharmaceutical industry. The percentages are an estimate of the current split of time between transforming and analyzing the data.

PhUSE 2009

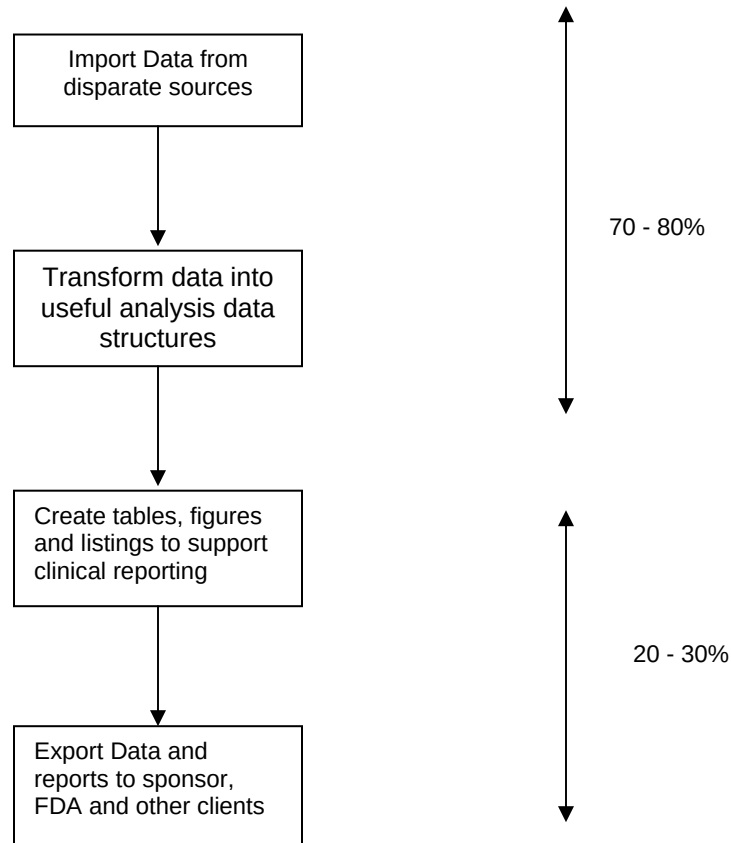


Figure 1: A simple illustration of the general work processes of the statistical programmer
"SAS Programming in the Pharmaceutical Industry" – Jack Shostak

So how have we reached this situation? Fundamentally, Clinical Trial Data are dirty. There are 3 main reasons for this:

1. Frequently, CRFs are designed without proper consideration to the analyses that will be performed on the data collected. If statisticians and/or programmers were involved earlier on in the process, the data would be more fit for purpose. This would mean less time is required to clean it, leaving more time available for the analysis of the data.
2. Data management fail to clean the clinical data properly prior to handing it to the SAS programmer for producing the TFLs. The development of electronic data capture (EDC) has resolved much of this problem, but it is not infallible.
3. Finally, we have to remember that we are dealing with human beings. Despite the best efforts of the clinical trial team, in developing a sound protocol and CRF, and the guidance of those running the trial, it is not possible to control the quality of the data that the patient provides.

This situation, of having to spend a large proportion of our time querying and cleaning data before analyzing it, is not unique to the Pharmaceutical industry. It can be seen across many industries where large volumes of data are used. Working with data from a range of clients, each with their own data storage/management systems can lead to a similar situation, with a programmer or analyst having to fix the data before it can be used. This is more of a problem in less regulated industries. Data remains the property of the client so many feel there is no need to 'fix' it – that remains the duty of the analyst or programmer. We see time to produce analyses squeezed further along the critical path. The analysis and interpretation of the output then has to be done very quickly and can result in poor quality output.

DO THINGS HAVE TO CHANGE?

The illustration above clearly shows an imbalance in the proportion of time spent by programmers spent on steps in the TFL production process. The question is: is this a bad thing? Before examining whether the current situation needs to change, we should consider whether things are fine as they are.

PhUSE 2009

Whilst the current balance is weighted in favour of data manipulation rather than analysis, one could argue that this results in programmers who understand the data well enough to know if the results of analyses are 'correct'. But is this the best use of resources? We are all under pressure to deliver clinical reports to tight deadlines. By having to focus so much attention on fixing data so that it is fit for purpose, we risk leaving insufficient time to produce and validate output. Many programmers have a range of skills beyond those required to validate data that are currently being under-used. This can lead to them feeling frustrated at not being able to use their full range of skills.

THE FUTURE'S BRIGHT, THE FUTURE'S CDISC

The pharmaceutical industry is undergoing a period of significant change. There is growing competition from generic drugs, increasingly tough P&R, and tighter regulation. Fewer drugs are gaining approval, and a decline in ROI means that the industry is going through a process of cost tightening. There have been significant job cuts and many companies outsourcing clinical trials in an attempt to optimize the R&D function. Whilst many of these changes won't have a significant impact on the day-to-day role of a SAS programmer within the pharma industry, the tighter regulation that is evolving, and the resulting impact on standard operating procedures almost certainly will. Regulatory authorities play an important part in the work of a SAS programmer. ICH-GCP, CDISC and FDA regulations all have to be adhered to.

CDISC is the biggest regulatory change that has happened in recent years in the Pharmaceutical industry. As it is likely to become a mandate by the FDA, its impact is going to be huge, and it is vital that as an industry we embrace it fully. Companies that have implemented CDISC standards have seen significant reductions in time and costs in areas such as development and review of CRFs, building and testing databases, and working with CROs.

One of the core principles of CDISC is that of the development of standards to improve process efficiency. Its recommendations are made up of various data models, a number of which will directly impact the work of the SAS programmer.

The Study Data Tabulation Model (SDTM) designs study data that is formatted to fit very specific data standards. The datasets are designed for submission to the FDA, so they should follow the same structure, even across studies. Clinical Data Acquisition Standards Harmonization (CDASH) focuses on developing standard data collection instruments. Analysis Dataset Model (ADaM) provides standards for creating analysis datasets. It contains the variables from SDTM, and adds all of the derived variables required to perform the analyses.

Whilst we will never be able to remove the element of human error from data collection, the development, and more importantly, implementation of CDISC standards (particularly SDTM and CDASH), should ensure that the quality and structure of clinical data will be embedded up-front at the start of clinical reporting process rather than being an afterthought. It should streamline data collection, resulting in the programmer spending less time on data handling and more time on the analysis of data and production of TFLs. We can expect to see higher quality output as a result and programmers becoming more efficient.

STATUS QUO- WHAT HAPPENS IF THINGS DON'T CHANGE

Whether we like it or not, the role of a programmer within the pharmaceutical industry is currently changing. It is not possible for us to remain unaffected by those changes. The implementation of CDISC and the resulting standardization of data will be the driver for this change. If we are not prepared for this change, and will to develop, we will find ourselves left behind.

CONCLUSION

In conclusion, the clinical reporting process is undergoing a period of substantial change which will affect everyone – programmers included. The roll-out of CDISC and the resulting standardization of clinical data should, in theory, make our lives easier. With studies designed to fit a pre-defined standard from the outset, there will be less need to spend time fixing and cleaning data.

Programmers are going to have become fully versed in CDISC and the aspects of it that affect their role. This should see a shift in focus, with the current ratio of 80:20 in favour of data manipulation moving towards more time spent on analysis. If programmers don't already have analytical/statistical skills, they are going to need to develop them. We may also find ourselves relying less on SAS, and more on other software solutions, as companies develop their own packages to implement CDISC.

During this period of change, there is going to be a lot to learn, and the traditional role of the programmer within the Pharmaceutical industry will be redefined, and hopefully for the better.

PhUSE 2009

DISCLAIMER

Please note that the opinions expressed in this paper are those of the author and do not necessarily reflect those of Quanticate.

REFERENCES

- (1) www.cdisc.org
- (2) "The Pharmaceutical Industry – Current and Futures Trends and Strategic Issues shaping Pharma" - *Datamonitor*
- (3) "SAS Programming in the Pharmaceutical Industry" – *Jack Shostak*
- (4) "Backpacking your way through CDISC: A Budget-Minded Guide to Basic Concepts and Implementations" *Brown and Fulton, SAS Global Forum 2009*

CONTACT INFORMATION

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

Gemma Bishop
Quanticate
Ground Floor, ATS Aviation Centre
Columbus Avenue
Manston Business Park
Manston
Kent
CT12 5DD
United Kingdom
01843 824586
gemma.bishop@quanticate.com

Brand and product names are trademarks of their respective companies.