

The expanding role of the statistical programmer – our experience

Punitha Philip, Cmed Clinical Services, Horsham, United Kingdom
Minola Ilica, Cmed Clinical Services, Timisoara, Romania

ABSTRACT

The role of the statistical programmer has been rapidly expanding over the past few years and the poster explores the skills accumulated by the statistical programmer to meet the different facets of the statistical programmer role. Though programming and production of tables, listings and figures still remains at the heart of the statistical programmer role, project management, data visualization and the use of multiple software tools are also equally important. By expanding on the different responsibilities of the statistical programmer, the poster highlights the need for programmers to constantly update their existing skill and learn new skills to develop along with their expanding role.

INTRODUCTION

The role of “The Statistical Programmer” has changed quite significantly over the past few years with different responsibilities and skills sets being required by the role. By expanding on the different facets of statistical programming, the poster highlights the need for programmers to constantly update their existing skills and learn new skills to develop along with their expanding role.

PROGRAMMING OF OUTPUTS

TABLES, LISTINGS AND FIGURES

TFL programming in essence hasn't changed very much over the years, however, developments in software and infrastructure have allowed for a more efficient way of doing things. As software packages offer more complex functions, a few of the formerly daunting tasks (like for example pagination) are now taking significantly less time and effort. However, programmers need to keep up-to-date on the new functionalities available so that they can be effectively utilised.

QC of outputs has become increasingly automated with electronic comparison gradually attempting to replace manual checking and output ticking. As a result, there is an increased scope for the programmer to not only use their programming skills but also use their knowledge of other software packages to enable quicker and more efficient QC of outputs.

DATA CLEANING REPORTS

Increased complexity of study designs has resulted in more outputs being required to allow correct study management, for example event tracking, listings of RECIST, listings to check unit conversions, listings to identify protocol deviations. These need to be easily accessible by other study team members. In order to do this the programmer needs to understand the rules required in order to produce the output and also needs to have knowledge of the required software tools to ensure the sharing of the outputs in a timely manner.

CHECKS ON DATA

Data checks have always been done to make sure that the data received was free of errors and fit for purpose, there is a responsibility on the programmer to question whether the data received is correct. As a result, it is very important that the programmer has in-depth understanding of the data they are working with and the issues affecting it. They also need to be aware of methods to display the data in the most effective way to aid the understanding of it. This could include subject level listings, summaries and figures. Also, regardless of how robust programming is, it is implied that a certain number of assumptions are being made on how the data are expected to look. In order for code to work properly even in unexpected cases, the programmer be able to firstly design code that works on as few assumptions as possible. If assumptions can't be avoided, they should be clearly documented and robust checks should be added to ensure that the data is compliant. The programmer requires good algorithm design and defensive programming skills to program assumptions and the checks for compliance.

RESTRUCTURING STUDY DATA TO MEET SPECIFIC STANDARDS

With the development of standard data structures, the programmer's goal is no longer only the programming and delivery of outputs, but also mapping of the data to the standard structure. As a result an in-depth knowledge of the standards is very important for programmers.

PhUSE 2015

As standard data structures are used more prevalently, the programming of data to these structures also need to be increasingly standardized. Programs would need to be written in a way to be able to cope with different structures in the input data to increase efficiency. So the programmer would need to be able to create robust code that can be easily adapted across different projects and for different clients, thereby allowing for shorter delivery timelines.

PROJECT MANAGEMENT

The relatively recent events in the financial world have created a sense of budget awareness within the industry, making project management skills a part of the Statistical programming role.

Even though senior managers have a top level oversight, the programmer is expected to provide an estimation on how long each delivery takes to complete and when the resources are needed in order to meet the timelines for each individual project. Apart from timelines, when looking at the resource requirements, the lead programmer will need to take into consideration the appropriate distribution of expertise within the team and would also need to make sure that the resource requirement are in line with the contract and the given budget. The programmers also need to coordinate the day to day activities on the project and keep track of status across different projects making use of the tracking tools available. So programmers needs to hone their resource estimation, resource management and budget management skills for effective project management.

COMMUNICATION

During the past years there have been significant changes to the ways team members communicate. As a result of massive advances in internet security and computer infrastructure, face to face interaction is no longer mandatory, reducing the need for travel enabling programmers to work from home. As meetings can now be held in a virtual environment, instant messaging (IM) has become an important feature in informal communication, gradually replacing the telephone conversations in situations where urgent response is required. The programmer has to update their communication skills to use these new advances in technology, be able to work with multi-cultural teams to effectively coordinate tasks across different time zones and countries and build good work relationships in a virtual environment.

IMPROVEMENT OF EXISTING PRACTICES

FINDING INNOVATIVE WAYS

Identifying areas and processes that can be improved and finding new and innovative ways of increasing speed and efficiency of programming have always been a part of statistical programmer role. But programmer's expertise is not restricted in identifying areas for improvements within their projects as their skills can be used in other areas such as analysing data collected by the business and checking budgets and metrics.

STANDARD MACROS AND TEMPLATES

Identification of the need for standard macros and templates to increase efficiency of programming is and always has been part of the role of the programmer. But as the process improvement work needs to be balanced with day to day projects, the programmer needs to pay increasing attention to planning their work and that of the teams requiring good project management skills.

DEVELOPING QUICKER AND MORE EFFICIENT METHODS OF PROGRAMMING AND VALIDATION

The methods of QC and validation are moving away from manual checks to explore methods of electronic validation, also allowing for a more risk based approach, requiring strong decision making skills as more decisions are left to the discretion of the programmer.

There is more emphasis on understanding the data and making sure that the programs are set up in such a way to cope with potential data inconsistencies. The programmer needs to make sure that they make use of standard code where possible to minimise the amount of checks required as standard code would have been already validated. The programmer needs to be attentive to detail and spot issues with data and programming code while setting up the initial programs, to preempt additional rounds of QC.

As validation includes not only checking of output data, but also checking of the whole output (layout, spacing, etc.) and having an oversight of the whole deliverable, the programmer needs to also make sure there is consistency between different outputs and previous deliverables. This necessitates that the programmer has a good overview of the project to ensure consistency and good documentation skills so that the history of the project and deliverables are well documented.

DOCUMENTATION AND TRACEABILITY

TRACEABILITY

Over the years there has been a significant increase in QC steps and checks due to the increased complexity of study designs particularly the volume of studies of an exploratory nature. It is imperative that every transformation is traceable, and corresponding documentation created. As more steps have been introduced over the years ensuring

PhUSE 2015

traceability requires additional planning and organisational skills than before. The programmer also needs to have good documentation management skills to make sure that these additional steps are thoroughly documented to ensure reproducibility.

DOCUMENTATION

As a result of electronic storage being less expensive and as a large part of study documentation is no longer restricted to paper, the industry is increasingly utilising electronic filing and signatures. So the programmer needs to develop good documentation skills to be able to manage increasing volumes of documentation and be familiar with the different document management systems used to effectively store and manage project documentation.

USE OF DIFFERENT PROGRAMMING LANGUAGES

PROGRAMMING LANGUAGES

SAS® has always been the gold standard in the industry. Alternatives like R® and Python®, have only been recently considered, for increased flexibility to apply to advanced analytics, interrogative visualization and links to other systems. Due to being an open source and being extensively used in academic environments, both R and Python are a quickly evolving tool with growing functionality.

As clinical statistical programming is not always restricted for people with programming backgrounds, an additional consideration with R is a low level language requiring a better understanding of programming languages and coding in general, while SAS is fairly easy to learn and user friendly, particularly if you have already been exposed to data manipulation.

Statistical programming has in many cases meant SAS programming. However, as other languages gain popularity the programmers would need to expand their skill set and learn more programming languages.

Also working with more dynamic programming languages will require that the programmer is able to come up with better validation techniques to test new features available that have not yet been validated. So the programmers also require good validation skills, so that they can successfully validate programming packages.

OPERATING SYSTEMS

As the server and cloud based infrastructure gains more popularity all around the world, statistical programming departments are also switching from working on PC based software, as server and cloud based systems could provide better and more stable working environment at a decreased cost. Switching to new operating systems would mean the programmers would be involved in migration of projects and hence would need to be acquainted with how programming software and software tools behave on different systems.

VERSION CONTROL SOFTWARE

Traceability has always been a key concept within clinical statistical programming. Version control software is increasingly used to ensure traceability of programming changes, as it offers a clear audit trail to all previous versions of any electronic file. Different version control software are used across the industry and the programmer needs learn about the different systems available so that they can make good use of the products available.

CONCLUSION

Though programming still remains at the heart of the programming role, the constant search for more efficient alternatives to the traditional techniques have highlighted the need for programmers to develop additional skills such as good algorithm design and defensive programming. Keeping up-to-date with different software functionalities, packages and tools is also essential, as the programmer needs to be able to easily adapt their knowledge to new programming languages and environments.

In order to effectively manage resource and budget, project management and tracking skills are also required, while good communication and documentation skills have also become an integral part of the programming role. Therefore it is important that programmers constantly update their capabilities to keep up with the changing requirements in the industry.

CONTACT INFORMATION

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

Punitha Philip
Cmed Clinical Services
Registered: Cmed (Clinical Research Services) Limited (3859092)
Holmwood, Broadlands Business Campus
Langhurstwood Road

PhUSE 2015

Horsham, West Sussex, RH12 4QP, United Kingdom
Work Phone: +44 (0)1403 755636
Fax: +44 (0)1403 755051
Email: pphilip@cmedresearch.com
Web: www.cmedresearch.com

Minola Ilica
Cmed Clinical Services
Cmed SRL
Str. Coriolan Brediceanu, Nr. 10
City Business Centre Building A, Second Floor
Timisoara, Timis, 300011, Romania
Work Phone: +40 (0)356 731 007
Fax: +40 (0)356 004 364
Email: ilica@cmedresearch.com
Web: www.cmedresearch.com

Brand and product names are trademarks of their respective companies.