

From data management to statistical programming: a real life change

Benjamin Lerbour, Novartis, Rueil-Malmaison, France

The opinions expressed in this presentation and on the following slides are solely those of the presenter and not necessarily those of Novartis. Novartis does not guarantee the accuracy or reliability of the information provided herein.

ABSTRACT

From entering into a professional life through data management to being in charge of statistical programming activities on several clinical trials, it takes a lot of trainings, adaptability and learning from mentors. Some important moments and steps should not be skipped to enter in this world of coders.

Once the transition is done, it is not always easy to act in this new role and stick to it under certain circumstances. On the other hand it is also possible to use your new skills to help your previous role: a programming procedure can save hours of data management review! Some tips will be given to combine the experience obtained in both roles and get the most from them.

INTRODUCTION

Data management and statistical analyses in the pharmaceutical industry are closely linked but the skills required from a person to act as a data manager or as a statistical programmer are very different. The intent of the first part of this paper is to explain the different steps required to move from data management to statistical programming and more generally to adapt to a new role. The 2nd part will give hands on examples of ways to combine skills acquired in both roles. Finally, a personal opinion will be given on the benefits of working as a statistical programmer, did the dream come true?

SUMMARY

Background

Background for the transition

This move from data management to statistical programming was a succession of a few steps. After graduating from a Master in computing and biology, I started in data management for one year through 2 internships and a short term contract. It gave me an idea of what this role was in a large pharmaceutical company and that it did not allow me to use my computing skills. Novartis was very helpful in facilitating the transition for me and I started by one year as a contractor before being hired internally as a statistical programmer.

Moving from DM to statistical programming

Main similarities and differences between the 2 roles

Data management and statistical analyses are closely linked. When working as a data manager, a part of the data checks you run are there to allow accurate statistical analyses. And as a statistical programmer, you interact with the data review team to help in getting cleaner data.

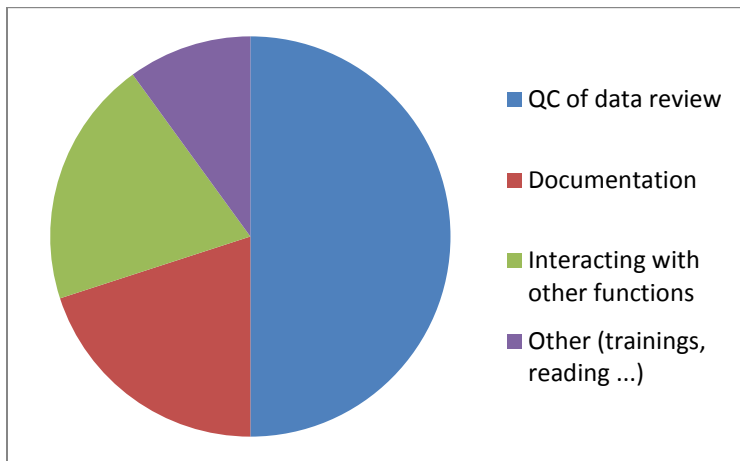
Similarities exist at project management level: a data manager works with CROs or with other data managers and ensures alignment of these teams to get complete and clean data. The programmer also works with a team to ensure validation of statistical outputs. Even though the interactions within the clinical team are different in the 2 roles, the team is the same and both roles work in the same direction.

Differences are more obvious in the day to day technical activities: the programmer focuses on statistical analyses and coding. The data manager focuses on ensuring data quality according to the data conventions he follows.

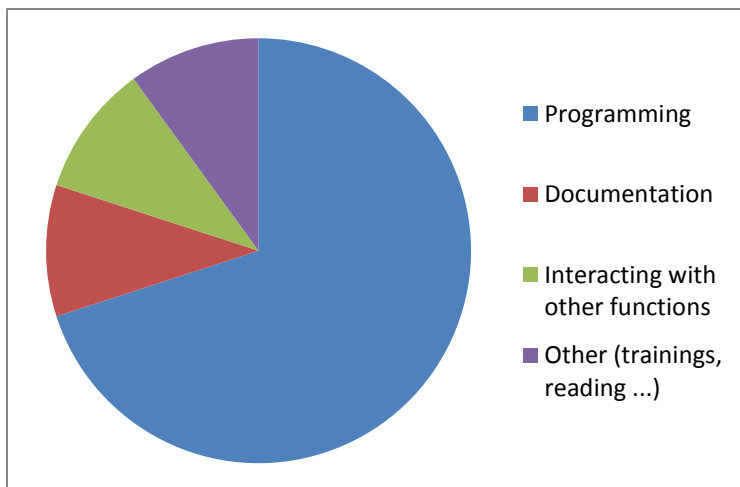
Here are diagrams of time repartition from the 2 roles, based on my experience. 2 diagrams represent programming roles as the time repartition changes a lot once you lead a team of programmers.

PhUSE 2015

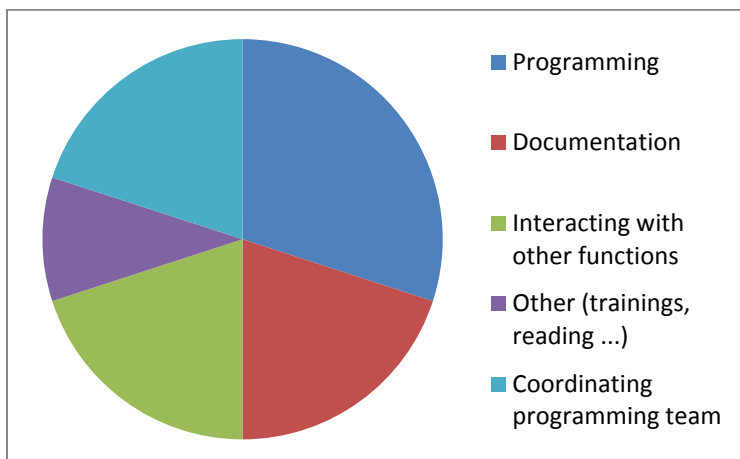
As a trainee data manager:



As a support programmer:



As a trial programmer:



PhUSE 2015

Reasons for moving from a role to another

For someone with computing and biology skills, data management with a sponsor gives a very good overview of how clinical trials are conducted but it can become frustrating not to be able to make profit of programming skills. Now was it possible, with this knowledge of clinical trials and computing skills, to program statistical models and analyze clinical data?

Acquiring necessary skills to be a statistical programmer

Currently, most junior statistical programmers in France and Switzerland come with a statistical background or at least statistical programming education. However it is not an impossible obstacle to start without this background if you show a good sense of adaptability and follow some important steps.

Starting with SAS trainings is obviously a necessary step to acquire confidence when programming as per a statistical analysis plan and providing the expected output. It is even more profitable if you also take a deep look into existing programs of your company, to understand how standard outputs are standardly developed.

Then to overcome the lack of understanding of complex statistical models, it is quite complicated to acquire knowledge that other programmers got in several years of school learning. But most of the time, a colleague will have the answer you are looking for. In the end, it is not uncommon that a programmer has questions on statistics and that biostatisticians need to explain.

Finally regarding the role of the statistical programmer in the clinical trial team, it is important to observe how your project leader is acting and learn from that. Even if you are a junior programmer, ask questions about how senior programmers interact with the team.

Experience sharing

Important situations

Quickly after moving to statistical programming, it becomes clear that the majority of your assignments have changed, but a statistical programmer may face the same situation as a data manager.

In both roles, it occurs that clinicians ask you for tools to review data. And in both roles, these requests come in addition to your priorities. In my experience of these situations, it is quicker and easier to provide these tools (listings, filtered tables) as a statistical programmer but it is also more out your scope. It is important in these cases to help your team but also to make them understand that statistical outputs are not data review tools and that you cannot program non-validated data review tools.

For more complex issues such as merged data (e.g. PK) or data trends (e.g. patients with static lab parameters between visits) that may bias the analysis, a statistical programmer can create some checks and alert DM if he finds anything abnormal.

Here is an example of simple code, in the middle of a PK merge, to identify and report potential data issues:

```
*-----  
| Bring PHK and DMG together and reconcile data |  
*-----;  
data pkdmg pkonly dmgonly;  
  merge  phk1 (in=pk )  
        dm1 (in=ind) ;  
  by &dmgsort;  
  
  if ind and pk then output pkdmg;  
  if pk and not ind then do;  
    Put "WARN ING: Data issue 1: PK records not in clinical DB Record number=" _n_ stysid1a= smp1n= phkrs1ln=;  
    output pkonly;  
  end;  
  if (not pk and ind) or (ind and smpsta1a eq "Not Received") then do;  
    Put "WARN ING: potential Data issue 2: subject no PK records and need to verify. Record number=" _n_ stysid1a= ;  
    output dmgonly;  
  end;  
run;
```

In many other cases of creating derived datasets, some simple code lines can be included into the SAS processing to ensure timely identification of data issues. It will save time for DM and also avoid any late issues preventing from creating datasets used for statistical analyses.

PhUSE 2015

Combining experiences

It happens for statistical programmers to be involved in data management processes. During the study set-up or when approaching database lock, your input may be required. I learnt that it can save time to be highly involved in the edit check review. These data checks are automatic. If they are well specified and take statistical analysis concerns into account, they will avoid manual checks for the statistical team and manual queries for the data manager. Here is an example of an edit check that was missing in specifications, but needed to be added to avoid issues in the interpretation of dose changes:

CRF Page/Form #	Evaluable Question/ CRF Item	Discrepancy Condition to generate discrepancy	Query text
TREAT_% (corresponds to any Treatment cover page), PRESC	LELCHG1C, SMDSTT10	LELCHG1C = No at Visit N and VIS10 at visit N < SMDSTT10 < VIS10 at visit N+1	The dose level change is "No" but a new dose prescription is reported. Please update.

In the majority of your new assignments as a statistical programmer, the skills you will use are those related to statistics and programming. But there are several situations when you can benefit from your knowledge of clinical data and help in improving data quality.

Your data manager may not have all the tools he needs for data review and if you provide him with a visual report, he can save time by avoiding manual filters or manual review. An example of lab outliers is given in the poster.

Finally, this a list of clichés data managers and statistical programmers can have on each other. Based on my experience from the 2 roles, I have answered:

	Cliché	Reality
Data manager	"looking forward to the test run of outputs to complete my data cleaning"	They are often made of dummy data and the purpose is to verify the layout.
Programmer	"I have spotted this issue a hundred times, why isn't it resolved?"	Sites can have different reactions to queries. Sometimes they will confirm values that do not look correct to the programmer.
Data manager	"it shouldn't take them long...they just have to press a button."	Unfortunately not, each study is specific and programs are adapted.
Programmer	"it's bad quality data...as usual."	A precise review of edit checks allows better data quality throughout the conduct of the study and reduces manual review.

CONCLUSION

This move is not standard, generally you « are born » statistical programmer or data manager. There is no standard transition package.

But nothing between the 2 roles is unbridgeable. A good sense of adaptability is primordial to overcome role differences. People will not mind your lack of biostatistics knowledge if you show other qualities.

After acquiring confidence in the statistical programming field, the combination of experiences is possible and it can save time to the whole clinical team when it comes to data cleaning specifications and data review tools. It also helps yourself understanding data you analyze and how it is structured.

It was a challenge for me at the beginning to enter the world of SAS programmers but the transition went even better than I had expected. This kind of life change, despite the efforts that it takes, is possible and profitable.

PhUSE 2015

ACKNOWLEDGMENTS

Thanks to Jerome Lechere, senior statistical programmer, for sharing with me some phrases he heard from data managers on statistical programmers, and vice-versa.

RECOMMENDED READING

It is recommended to read job descriptions of clinical data manager and statistical programmer roles under Novartis careers website:

<http://www.novartis.com/careers/index.shtml>

CONTACT INFORMATION

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

Benjamin Lerbour
Senior Statistical programmer
Novartis Pharma S.A.S.
10, rue Lionel Terray
92500 Rueil-Malmaison
FRANCE

Work Phone: +33 155476546

Email: benjamin.lerbour@novartis.com

Web: www.novartis.com

Brand and product names are trademarks of their respective companies.