Paper PD08

Managing the Change – Evolving from Statistical Programmers to Clinical Data Scientists

Sascha Ahrweiler, Bayer Healthcare, Wuppertal, Germany

ABSTRACT

Working as a statistical programmer in the pharmaceutical industry comes along with working in a rapidly changing environment. Frequently updated regulations, cost efficient outsourcing opportunities, and the evolution of standards are just some of the challenges pharmaceutical companies are currently facing. Such challenges should be perceived as positive that come with new opportunities. Statistical programmers often have the right mix of an analytical mindset and thorough understanding of data structures and contents. Statistical programmers employed at pharmaceutical companies may have the opportunity to evolve to new roles and acquire a broader understanding of data as an asset for the company. Advanced analytical skills and tools will be required for the next evolutionary step as well. This paper outlines how leadership can support the development of statistical programmers towards a new strategic role which might be termed Clinical Data Scientists.

EVOLUTION OF STATISTICAL PROGRAMMER

Over the past four decades, the evolutionary phylogenesis of the statistical programmer has developed from a “caveman” era to a highly advanced and skilled species in order to be successful in the changing and evolving world of drug development. In the early 20th century, randomized clinical trials became embodied in regulation as government authorities began recognizing the need for controlling medical therapies. Along with scientific advances, the need for advanced statistical analysis increased rapidly and could not be managed solely by Biostatisticians. In the late 1960s, the role of a Medical Documentalist was created to support the Biostatisticians with the preparation of tables, figures and listings for clinical study reports. They often used batch-computer programs submitted to large mainframe-based computing centers and later used programs, which were developed by the Biostatisticians in such programming languages like Fortran, PL1, or even BASIC. In this pre-personal computer era, a lot of manual work also included the drawing of summary profiles to visualize clinical study results. Later, the manual drawing was replaced by literal “cut-and-paste” of plots printed on plotters into clinical trial reports. This was the earliest era for statistical computing, and the beginning of the evolution.

With the breakthrough of personal computers in the early 1980s - and especially with the release and continuous development of SAS® software in the late 1970s and early 1980s (historical note: SAS79 was written in PL1 for PROCs and Assembler for the DATA step1) - the role of the Medical Documentalist changed drastically and we evolved out of the “caveman” era to an era of juvenile development of statistical programming. The interest and usage of new statistical models, such as generalized linear models and the ability to analyze such models - with SAS® through a user interface - increased with the availability of computer power. This increased need came along with the need of new experts who could develop SAS® programs and, with that, the evolution of the Medical Documentalist to the SAS® Programmer.

Initially, responsibilities of SAS® Programmers were clearly separated from those of statisticians. Clinical Trial Statisticians developed more sophisticated portions (such as the analysis of efficacy using analysis of variance), while SAS® Programmers took care of data handling or summary statistics. Over time and with increased regulatory needs, more pharmaceutical companies made a clearer distinction between these two functions. Clinical Trial Statisticians focused solely on the planning of clinical trials, while the SAS® programmers implemented these plans and translated statistical analysis plans into SAS® code. Along with an increased need of statistical knowledge among SAS® programmers, this role further evolved to the *Statistical Programmer*.

NEXT STEP IN THE EVOLUTION

In retrospect, the evolution of the statistical programming role leads to the conclusion that statistical programmers need to continuously acquire new skill sets to keep up to speed with a rapidly changing environment. This continuous development process needs to be actively supported by line managers, who are also developing and enhancing their own skills to keep pace with the evolving environment. For any manager involved in personal development, it is important to have an idea about future trends and opportunities to ensure that staff acquires new skills and develops in the right direction. There is, of course, some uncertainty around how the entire pharmaceutical industry will develop, but in recent years, it has become very obvious that biometric functions need to think beyond submissions and incorporate complete life cycle management of a compound or device.

Until the early 2000s, it was sufficient that pharmaceutical companies prove the safety and efficacy in order to obtain marketing approval. After drug approval, companies could more or less define the reimbursement prices. Economic changes and shortages in global medical reimbursement budgets have led to some countries employing more stringent reimbursement policies. Today, instances like the German IQWiG, the British NICE, and US Private Health Insurance Companies assess new approved drugs for their cost-benefit ratio and recommend reimbursement prices. These evidence-based post-approval assessments require pharmaceutical companies to invest more time in researching unmet clinical needs significantly earlier in the development process, as well as incorporating these effectiveness endpoints into clinical trials for market access after approval. Thus, the phylogenesis of the statistical programmer has taken us to this stage in order to be successful, contributing members of drug development teams.

The amount of available clinical trial data (and beyond) has grown exponentially over the past decade, with data coming from a myriad of sources:

- In various countries, linked healthcare givers share observational patient data to enable many types of research. For example, the Clinical Practice Research Datalink (CPRD) [3] is an English NHS observational data and interventional research service, jointly funded by the NHS National Institute for Health Research (NIHR) and the Medicines and Healthcare products Regulatory Agency (MHRA). Via collaborations, CPRD enables access to data of more than 20 million anonymised patient lives.
- Today's patients use smartphones or internet applications to track their daily disease symptoms. This data might be shared with physicians for optimal care, but also with other patients to learn more about the disease. Patient portals like PatientsLikeMe [4] allow patients to share their data to support other patients and research. Since the inception of PatiensLikeMe in 2004 more than 250.000 patients have reported more than 7 million reports of their treatments and symptoms and more than 50 research studies were supported.
- Pharmaceutical companies have to become more transparent and release data collected in clinical trials. The publication of the EMA policy [5] on release of clinical trial data requires pharmaceutical companies to share their anonymized data from clinical studies to allow research.
- Patient registries [6], a collection of information about individuals, usually focused around a specific diagnosis or condition, are built up on a voluntary basis. These registries can be sponsored by a government agency, nonprofit organization or health care facilities. Also private pharmaceutical companies might need to commit to sponsor registries as part of their post-marketing requirements. These Patient registries help especially patients with rare diseases advance scientific knowledge and improve patient care.

There are countless other examples which have the common theme - there is a huge amount of disease and treatment outcomes data available in the healthcare industry. The connection of various external and internal data sources, and its analysis and conclusion might eventually lead to a more efficient drug development process. The understanding of data and its correct usage will distinguish between the success or failure of companies, and it will also play a key role in the next evolution step for statistical programmers.

If one would like to give the next evolution step for statistical programmers a name, one could call them *Clinical Data Scientist*. While there is criticism in other industries about the usage of the term *Data Science*, it will be used for the remainder of this paper to distinguish between current and potential future roles. The term Clinical Data Scientist will also be used to summarize the required skill sets necessary for this future role.

MANAGE THE CHANGE TO THE SKILL SETS REQUIRED

The required skill sets for a Clinical Data Scientist are defined by the scope of work for a Clinical Data Scientist. On their webpage, the PhUSE community[7] presents a generic job description for a Clinical Data Scientist. Clinical Data Scientists are described with the following eight keywords:

#1. Clinical Trial Expertise	#2. Data Integration
#3. Scientific Rigor	#5. Collaborative
#5. Statistics	#6. Advanced Computing
#7. Visualization	#8. Hacker Mindset

This definition of Clinical Data Science shows that especially Statistical Programmers have the right mix of skills. They are ready to make the next evolutionary step. Depending on the actual tasks, further development of certain soft or hard skills will be needed. For example, a Clinical Data Scientist who works on clinical trials following exact specifications from a Clinical Trial Statistician will need less statistical expertise than a Clinical Data Scientist who works on exploratory tasks without any involvement of a Statistician.

Although Statistical Programmers have the right mix of skills to quickly adapt, the evolution towards the Clinical Data Scientist will still require change, with a steep learning curve. With this, the basics of change management need to be kept in mind as well. There needs to be a clear commitment (both time and financial) from the upper management on the strategy to evolve the knowledge of statistical programmers. The biometric leadership needs to develop a clear vision and a roadmap to help the entire staff understand the upcoming challenges and opportunities. The vision and strategy need to be communicated frequently in order to ensure that all personnel receive the correct message. However, not everybody will adopt at the same speed. There will be fast adopters and slow adopters. For a successful change it is of critical importance to identify the early adopters and utilize them to convey the message. Finally, one must be aware that this evolution will not happen overnight and will take some time. It will be a journey, which will be more enjoyable if the entire team is looking forward to reach the destination.

In the following sections, the needed skills will be explained in more detail together with ideas how managers can help to further develop skills and manage the change. All examples have a common theme that they will lead people slightly out of their comfort zones. While this might feel challenging in the beginning, moving out of a comfort zone is an absolute necessity for continuous development. The emphasis should however be on “slightly” and should build on an already existing matter expertise for which there is room for advancement. Even for Subject Matter Experts, it still requires personal engagement, openness and also trust to go through this change. As Statistical Programmers often have to cope with tight timelines in combination with a large workload, all examples chosen are considered to be time- and workload-efficient.

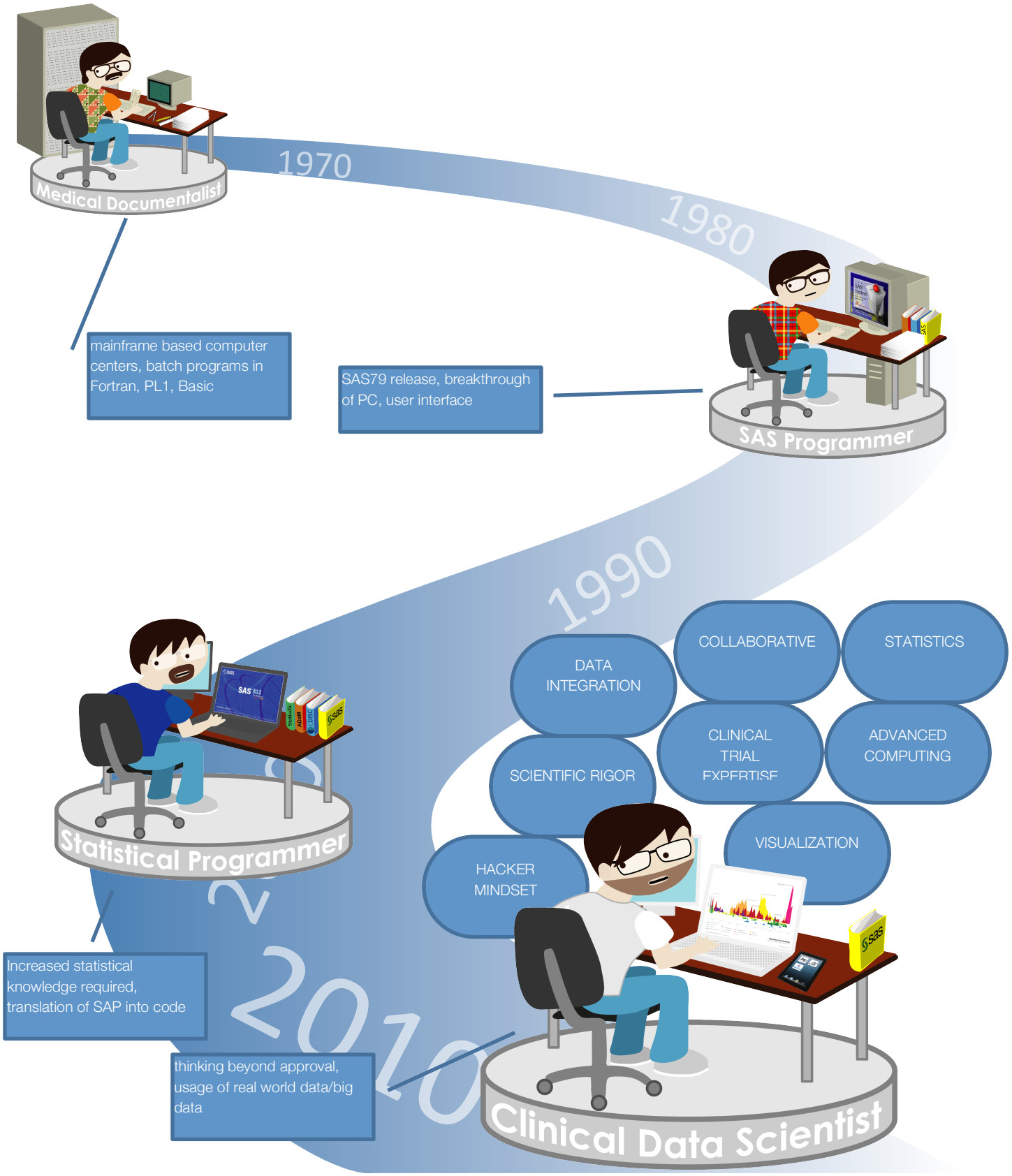
SKILL #1 - CLINICAL TRIAL EXPERTISE

A Clinical Data Scientist needs to understand the overall drug development process from research to drug approval (ie. Phase I to III). This includes a detailed understanding of trial designs, data collection, data analysis and reporting of clinical trials. Also, increasing CDISC SDTM and ADaM usage in clinical trials require expert knowledge of these data standards.

Clinical trial expertise naturally comes with years of experience as a Statistical Programmer. Given the huge amount of different study designs, it could take a long time before someone gains expertise in various phases, designs or indications. If individuals share their knowledge amongst their colleagues, this duration will be shortened.

A good way to share knowledge on clinical trials is to ask the study Lead Statistical Programmer (LSP), who co-ordinates all statistical programming efforts for a study, to prepare a presentation about the study results once the study is finalized. This presentation should inform attendees how the study fits into the overall drug development plan. It should also explain how the study was designed together with some justification of why the chosen study design is adequate for the research problem. Besides a brief walkthrough of the main results, also any study specialties should be presented. One example of a study specialty would be transparent data handling in a local affiliate study which could include solutions on CRF annotations and the consequences on controlled terminology. Often study specialties require adaption or even creation of existing standards and therefore it is more than helpful to spread this knowledge.

Evolution of the Clinical Data Scientist



Presenting study results to the entire statistical programming department has several advantages and they are time- and workload-efficient. Usually the LSP is involved in a very early stage in the study planning and set-up phase and will need to familiarize with most of the questions to be presented later anyway. Ideally, the LSP prepares a presentation for a kick-off meeting with the remaining statistical programming team members and just needs to slightly adapt this presentation after finalization.

We must keep in mind that the main objective here is on knowledge sharing and personal development. The study result presentation should only be used internally and therefore formal language is not required. While the text in the study protocol or clinical study report might be formally correct, the LSP should resist to ”cut-and paste” everything from these documents. The LSP should rather use their own words to describe the content; this will lead the LSP to increase the own awareness through discussions with medical or statistical team members, and also help the LSP to advance their own knowledge. This will naturally have a tremendous effect on teamwork and collaboration.

Finally, the structured storing of study result presentations will help to create a great knowledge library on drug development approaches and problem solutions. Based on the company size, someone might consider, for example, to create a central Wiki to store all this information. Using categories or keywords will be of critical importance to find information afterwards. This will help both experienced and new staff increase their knowledge.

While information sharing and knowledge increase is the main goal for the study result presentation, it might additionally be helpful to improve presentation skills. The presentation should be followed up with a feedback session of the presenter and the line manager. Here a videotaped presentation would also allow to discuss appearance and presentation skills very objectively and will help the presenter to realize progress over several years.

SKILL #2 - DATA PREPARATION AND INTEGRATION

Clinical Data Scientists work on integrated data analysis for various purposes such as integrated analysis of safety and efficacy (ISS, ISE), value dossiers for reimbursement purposes or health economic models. They need to know how to setup databases for various purposes or how to prepare data to fit various purposes.

One of the most important purposes to integrate data will still be the integration summary of safety and efficacy to receive a new drug admission from a regulatory agency. In a strongly regulated and standardized environment, it is of critical importance to fully understand the existing regulations and standards. Someone who needs to work on a CDISC compliant integrated analysis such as ISS or ISE needs to have intimidate knowledge about these standards. Attending official CDISC trainings or conferences such as PhUSE or CDISC Interchange events, will help to fully understand the standards and the issues which might come with the implementation. There is also no way around some reading of the CDISC implementation guides for SDTM and ADaM, as they contain guidance for the implementation. Reading documents is also a necessity when it comes to the requirements from the regulatory agencies. A Clinical Data Scientist should be aware of existing agency guidelines and should be able to follow them.

As stated in the sections above, increasing healthcare costs come along with an increased demand of additional health economic analysis of data. The data to be analyzed might include company owned clinical trial data, but it might also include external databases like the above-mentioned CPRD, and often also a connection of both internal and external data. Access to an external database is generally only granted after a formal training was taken. Such formal training will then help to understand the database structure and other formal aspects. In order to work efficiently, it will always be of critical importance to fully understand the data.

Understanding data will be a key skill for a Clinical Data Scientist and this can be achieved again through knowledge sharing as described in the previous section. In a clinical study setting, clinical study endpoints are assessed and need to be reported in a transparent and traceable manner. One way of understanding data is to follow the entire workflow of data starting from the patient assessment through databases and finally to the report. A structured way to share knowledge in a presentation is to choose a parameter of interest and to talk about the following topics:

Regulatory guidelines on a therapeutic area of the parameter

Data collection and preparation to become analyzable data (e.g. how is a X-ray image transferred into data points?)

Storage or remapping to SDTM data (e.g. does a domain already exist?)

Data derivations and storage in ADaM data (e.g. if a parameter is a composite score, how is this calculated?)

Statistics and data presentation (e.g. are there existing standard displays for result presentation?)

Points to consider for data integration (e.g. how to handle different methods or versions of data assessment and the impact on pooling strategies)

Any other challenges which might need to be taken into account for a certain parameter

Once a statistical programmer has gone through this entire data workflow more than once, it will become natural behavior to try to understand data in such a structured manner. A programmer will also rise the own awareness of disease areas and should be able to adjust this knowledge from clinical studies to other databases as well. It will also help a programmer to understand the necessary steps to be taken to reach a defined goal. This is a key skill for people who work on an integrated analysis which combine several studies to integrated datasets.

Understanding the goal of data integration or preparation tasks also includes the understanding of formal requirements of data. While there are strict regulatory requirements for electronic data submissions to regulatory agencies for a NDA, analysis tasks for scientific publications or health economic modeling or other non-regulatory relevant tasks are currently more flexible. It is of course possible to also implement data standards also on retrospective observational studies, but depending on the requirements it should be decided on a case by case basis taking the cost-benefit comparison into account.

Similar to the previous section, sharing knowledge through group presentations by subject matter experts (SMEs) come with additional benefits - through sharing and storing knowledge in a structured way, this will help to build a knowledge and standards library as well.

SKILL #3 - SCIENTIFIC RIGOR

Clinical Data Scientists analyze data and report results to regulatory or reimbursement agencies. The scientific rigor ensures a seamless and efficient dossier review and approval. As there are differences in the different therapeutic areas, Clinical Data Scientists need to understand that different therapeutic areas might require different approaches.

Scientific Rigor is a key skill for a Clinical Data Scientist. Most of the job descriptions for Statistical Programmers require a mathematical, technical or computer science background. As these are subjects which require accurate and precise working, such an education naturally trains a lot of scientific rigor by itself. In some cases, however, medical knowledge is not required, which requires some attention in personal development.

It has already been stated in the previous section, that therapeutic area knowledge is of critical importance for a clinical data scientist - there needs to be a thorough understanding of the challenges in a specific therapeutic area and the corresponding regulatory guidelines.

Besides the knowledge sharing of SMEs as explained above, another good source of knowledge is the investigators brochure (IB). This document needs to be updated regularly and is a requirement for every compound in clinical development. The IB contains comprehensive information about an investigational drug and is updated on a regular basis. It contains information about dosage, safety profile, methods of administration and other drug specific information. Reading and working with the IB helps the Statistical Programmer to become aware of challenges beyond formal data issues and is highly beneficial.

Another collaborative way to gain further insight into medical application is to go out in the field and accompany a study project manager at a site initiation visit. Especially in early drug development, where subjects are hospitalized for pharmacokinetic profiling, study nurses and investigators are trained on the study protocol and the pharmaceutical drug. Attendance at a site initiation visit helps a programmer to understand how the data is assessed and will lead to a different view on data. It helps to understand challenges in CRF handling and monitoring.

Once market access is granted, pharmaceutical companies also train their sales forces on the mode of action, safety profile and other specific information of their drug. Attending a sales force or medical affairs training therefore can help to better understand patients and treatment options as well. As most information in the drug label is a result of the ISS and ISE, it is also a good way to learn how the statistical programming work package is used.

SKILL #4 - COLLABORATIVE

Clinical Data Scientists closely collaborate with a lot of different functions inside and outside the biometric area. Within the biometric area, they closely collaborate with CRF designers, data managers, biostatisticians and pharmacokineticists. Outside the biometric area, they closely collaborate with other functions such as medical writers, clinical teams, project managers, health economist, publication managers, and - of course - patients.

Collaboration means working with each other to achieve a shared goal. It means sharing knowledge, learning from each other and building consensus. To reach a shared goal, it is not important how clever individuals are, it is important how smart the collective brain is.

Collaboration requires strong active communication skills. Learning from each other requires communicating to learn and not only to reply. Far too often, human beings are locked within their own way of thinking or understanding. A basic rule for communication therefore is, to probe that what you heard is what was said. Acquiring good communication skills is important for a Statistical Programmer who works on well specified tasks in a multifunctional team. It is of critical importance for a Clinical Data Scientist who works on less specified tasks though. Communication is a broad topic that also runs deep. Training courses can help to learn basic rules and they are especially helpful if they are customized and allow trainees to practice communication patterns in a workshop setting. But good communication needs intensive practice and also feedback from a mentor, coach or line manager.

For multifunctional teams, communication is the key to collaboration as it is the foundation upon which we learn to understand each other. For good team collaboration, all functions should be able to understand what other functions do in the clinical life cycle process, and why they are needed. Participating in short internships can ensure a deeper understanding of other functions. This can also be achieved by simple meet-and-greet meetings where main collaborators exchange their expectations. This exchange is helpful to put oneself into other function's shoes and helps to have more efficient and effective collaborations.

SKILL #5 - STATISTICS

Clinical Data Scientists should have a very good understanding of statistical methods used in clinical trials. They should also know how to apply statistical methods to deepen their understanding of data you work with.

In theory, Statistical Programmers should already have an understanding of statistical methods used in clinical trials and integrated analysis. However, statistical knowledge varies based on the education of programmers. A sound statistical understanding should be available with a mathematical background, while it may require some additional attention for programmers with other scientific background in other life sciences.

Statistical programmers who have provided tables, listings and figures for a clinical trial should be able to calculate standard statistical measures such as the mean or standard deviation. More sophisticated statistics such as mixed models or correlation analyses will require the understanding of more advanced statistical thinking. There are, of course, short courses available, but these often focus on students who already have a statistical background. For programmers without any statistical background, it might be easier to learn from statisticians who actively worked as programmers. A great source for articles is the PhUSE library of papers from annual conferences. They usually give a brief theoretical introduction related to a special statistical topic and help to understand programmatic solutions. But there are also some very good introductory papers like the paper Misinterpretation of Statistics – An Introduction from Salter8 which help to understand basic statistical thinking.

Mentoring by either colleagues from the biostatistics department or by SMEs within the programming team, is another way to gain statistical expertise; when a Statistical Analysis Plan needs to be reviewed by a programmer, it is highly beneficial if another programmer with good statistical knowledge probes the less experienced programmer, if everything was understood correctly. Here, the paper Training Statistical Programmers on SAP Review Skills from Ahrweiler 9 explains in more detail how a mentor could support a mentoree.

Of course, all the methods to share knowledge as described in the previous sections will also help to extend the statistical knowledge. If statistical SMEs share their knowledge by presenting statistical highlights in clinical trials, the non-statisticians in the team will benefit as well.

Regardless of the learning methods, the learning of statistical methods is quite complex in its nature. A Clinical Data Scientist does not need to have the exact same knowledge as a statistician. But he needs to be aware of his knowledge gaps and collaborate with other SMEs to minimise these gaps by asking the right questions.

SKILL #6 - ADVANCED COMPUTING

Clinical Data Scientists are able to use various tools for data analysis and visualization. They are not limited to a single tool and know which tool to use to analyze and report clinical data efficiently.

Someone who intentionally has chosen a job that requires working with a computer for at least 8 hours every day should have these advanced computing skills. However, it requires some openness to changes to leave secure paths and look into new tools. A good SAS® programmer will be able to solve the majority of programming challenges with SAS®. A great Clinical Data Scientist though knows that there are other tools available and that other tools might be able to deliver results quicker or might present them in a nicer fashion. The choice of the right tool also depends on the problem someone tries to solve. An agency submission will require other tools than an ad-hoc presentation of a data-mining task.

If a statistical programmer is open to learn about other tools besides SAS®, there are plenty of workshops, webinars or trainings available for various tools. Platform independent conferences like PhUSE also offer a lot of presentations about tools like Spotfire®, R, JReview, JMP Clinical, Qlik View or any other helpful tool for data handling and analysis. And someone should not forget about things he can learn from talking to software developers who exhibit at these conferences.

Again the creation of a knowledge library and the sharing of knowledge would be helpful. The main goal though should be the understanding of the pros and cons of applications to pick the best tool for a given task. This will help the Clinical Data Scientist to closely collaborate with an expert in the IT team, who should be the SME on questions related to the application management.

SKILL #7 - VISUALIZATION

Clinical Data Scientists know how to visualize all kind of data. They should have a good understanding of the creation and interpretation of informative graphs to summarize visually results for various purposes such as clinical trial results, integrated analysis or data mining.

Statistical programmers should have a good understanding of visualization methods used in clinical trials. They should at least be familiar with the creation of individual and mean plots, and bar charts. More advanced programmers are also familiar with scatter plots, boxplots, Kaplan-Meier curves or regression plots. There are other visualization methods like Forest and Funnel Plots for Meta-Analysis, or Tree Maps, which are often used in Safety Signal Detection analysis. And also the usage of Infographics to present analysis results in a graphical way is used more frequently.

There are countless books on the market for data visualization topics and several publications have been written as well. However, comprehensive examples on clinical data are rather limited. Therefore the development and creation of a knowledge repository might help to build the bridge between commonly used presentation methods and its application in medical data. A good source to learn about commonly used methods is a webpage, which shows the Periodic Elements of Visualization Methods10. This webpage should help to get a first understanding of available methods. A good exercise for a Statistical Programmer who wants to develop his visualization skills then could be to think about ideas how these methods can be applied for clinical data and share the lessons learned with other colleagues.

SKILL #8 - HACKER MINDSET

Clinical Data Scientists have advanced skills in problem solutions. Solving difficult data derivation challenges and finding appropriate solutions for every problem in an efficient manner drives them.

It might be prejudice towards people who are passionate about programming and who write algorithms day in and day out, but most of the statistical programmers have an innate curiosity to solve data problems. It is almost like solving a

puzzle, only that the puzzle pieces consist of data and their relationship. Finding out about this curiosity in a programmer is a method for leveraging their skills and development.

Among the skills needed for a Clinical Data Scientist, this "hacker mindset" is often the most difficult to train for a line manager. It requires people skills, time and personal interaction to find out what kind of puzzles still challenge a programmer. It especially requires a line manager to care about the staff development in addition to the interest of timely deliveries. This also includes that line managers should acknowledge situations when challenges were solved and do not take everything for granted. Acknowledgement of hard work and success tremendously helps to keep people engaged.

CONCLUSION

The role of the statistical programmer has evolved over the past and it will continue to evolve in the future. While a precise forecast of this future is almost impossible, this article outlined a potential next evolutionary step towards Clinical Data Science. With the right mix of skills, the Statistical Programmer should be able to advance in this direction. In times where companies need to think beyond submissions and need to understand data as an asset for the future, Clinical Data Scientists will help to create value by connecting data and create knowledge.

Even though this evolution could be considered as a slight evolutionary change, it is still a change that requires a steep learning curve and appropriate support from line management to help develop additional skills. As shown above, this can often be incorporated into day-to-day work as needed.

As long as people are open to change in our rapidly evolving environment and as long as they are engaged in their personal development and are willing to sometimes act out of their comfort zone, the examples provided above will help statistical programmers to evolve to the next level.

ACKNOWLEDGEMENT

The preparation of this article would have not been possible without the support of my previous line manager, mentor and friend Jim Johnson, PhD (Summit Clinical). The thought provoking discussions and especially the sharing of his insights into the "caveman" era of statistical programming were beneficial.

Thanks to my dear wife Telse Ahrweiler who supported me with the creation of the visuals in this paper.

I also would like to thank my colleagues John Salter, Brigitte Bernhoff and Gaby Jende, who supported the preparation of this paper with valuable discussions and comments.

REFERENCES

1. SAS (Software), Wikipedia entry, http://en.wikipedia.org/wiki/SAS_%28software%29, last checked August 22nd 2014
2. FDA Critical Path Initiative, http://www.fda.gov/ScienceResearch/SpecialTopics/CriticalPathInitiative/ucm076689.htm, last checked August 22nd 2014
3. The Clinical Practice Research Datalink (CPRD), http://www.cprd.com, last checked August 22nd 2014
4. Patients Like Me, http://www.patientslikeme.com/, last checked August 22nd 2014
5. Publication and access to clinical-trial data: an inclusive development process, http://www.ema.europa.eu/ema/index.jsp?curl=pages/special_topics/general/general_content_000556.jsp, last checked August 22nd 2014
6. NIH Clinical Research Trials and You, List of Registries, http://www.nih.gov/health/clinicaltrials/registries.htm, last checked August 22nd 2014
7. PhUSE, Data to Knowledge, http://www.phuse.eu/D2K.aspx, last checked August 22nd 2014
8. John Salter, *Misinterpretation of Statistics – An Introduction*, 2012 PhUSE annual conference proceedings, http://www.phusewiki.org/docs/2012/PAPERS/SP/SP02.pdf, last checked August 22nd 2014
9. Sascha Ahrweiler, Training Statistical Programmers on SAP Review Skills, 2011 PhUSE annual conference proceedings, http://www.phusewiki.org/docs/2011%20Papers/IS01%20paper.pdf, last checked August 22nd 2014

10. Periodic Elements of Visualization Methods, http://www.visual-literacy.org/periodic_table/periodic_table.html, last checked August 22nd 2014

CONTACT INFORMATION

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

Sascha Ahrweiler
Bayer Healthcare
Aprather Weg 18
42113 Wuppertal, Germany
Work Phone: +49-202-36-5323
Email: Sascha.Ahrweiler@bayer.com
Web: www.bayer.com