

## **Get kitted out – A practical guide to SAS certifications, MOOCs and hands-on experience**

Hannes Engberg Raeder, PRA Health Sciences, Mannheim, Germany

### **ABSTRACT**

This paper looks at three ways in which both beginning SAS® programmers and seasoned veterans can improve their skills and broaden their experience. One of the more obvious ways to improve SAS programming skills is to prepare for one or more of the SAS certifications available. Furthermore, MOOCs (Massive Open Online Course) have increased in diversity, popularity and availability in recent years, and due to the wide range of courses on offer they are also an interesting option for the inquisitive programmer. Possibilities and practical aspects are revealed here. Finally, one of the most effective ways of extending one's skillset is with hands-on experience. However, this can also be the most challenging way: it requires the programmer to step outside his or her comfort zone and their manager to recognize such development opportunities within the context of the company's business needs.

### **INTRODUCTION**

There are few, if no, established paths to expand your skill set and experience in the narrow field of statistical programming in the pharmaceutical industry working with clinical data. There are companies, both pharma and CRO, offering their own courses/programs for those entering the profession and for more seasoned programmers as well, nevertheless the vast majority will need to find their own route to build up experience and acquire new skills.

The paper does not claim to provide a solid solution for everyone, but strategies and concrete solutions that have worked so far for the author and others. In detail, the paper looks into four certifications currently offered by SAS that are of particular interest to the programmer in the pharmaceutical industry. Secondly, the concept of MOOCs is presented and the advantages for programmers are discussed. The last section presents different strategies to acquire hands-on experience as the means to reaching a certain goal.

These three aspects of building experience and developing new skills are not distinct. A programmer in the industry might very well become a SAS expert through hands-on experience only by working on different types of projects, testing new programming techniques etc., but going through the four certifications discussed below, might present the programmer with an accelerated path to achieve this goal of becoming a SAS expert. The same illustration applies to most other aspects of the programmer's daily tasks: Someone can become adept in understanding vital signs while working with data, asking medical writers and reading the study protocols etc., and here as well taking the right MOOCs might offer a fast track to a deeper understanding of certain aspects of the cardiovascular system. If the above examples illustrate the intersection between SAS certifications or certain MOOCs and learning by doing, there are skills that can easily, mostly at least, be acquired with hands-on experience. Becoming an expert in SDTM or ADaM is difficult only by reading the CDISC manuals and implementation guides, it's exposure to real life problems and data that's necessary, or becoming an expert in programming different types of complex table and graph outputs cannot be taught in courses or in a paper. The hands-on experience is required to develop these kinds of skills.

### **CERTIFICATIONS**

In the past, currently and for the foreseeable future it is safe to assume that SAS software will play a major role as the main software for data manipulation, statistics and reporting which means that investing time developing and strengthening your SAS skills will only be to your (and your company's) benefit. At the moment, summer 2016, SAS is offering a variety of SAS programs with associated certifications. Here the focus lies, for the clinical SAS programmer, on relevant certifications from the Foundation Tools Program, which are:

- SAS Certified Base Programmer for SAS 9
- SAS Certified Advanced Programmer for SAS 9
- SAS Certified Clinical Trials Programmer Using SAS 9

and the Advanced Analytics Program:

- SAS Certified Statistical Business Analyst Using SAS 9: Regression and Modeling

Here the four different certifications will be outlined and practical advice will be given on how to best prepare for the exams.

## PhUSE 2016

### **SAS CERTIFIED BASE PROGRAMMER FOR SAS 9**

As the name of the certification indicates, this is the fundament covering topics like accessing different data sources using various methods, creating data structures, generating reports using the PRINT procedure and handling errors. To get the latest information about the content, number of questions and pass mark, the reader is encouraged to consult the SAS webpage ([sas.com](http://sas.com) webpage). At the time of writing the exam consisted of 60-65 multiple-choice and short-answer questions for which the candidate must score at least 70% to pass.

The preparation for this exam will obviously depend heavily on the candidate's previous experience and current SAS exposure. SAS offers an official guide called *SAS Certification Prep Guide Base Programming for SAS 9* which covers the all topics in the exam syllabus. This should be enough for most of the candidates as it does cover all aspects of the exam (but not in that great detail). If the current job is not covering the necessary topics or you are new to SAS programming you might need to consult other literature as well.

SAS Functions – it is not unlikely that the exam contains questions about functions not covered in the *SAS Certification Prep Guide Base Programming for SAS 9*. SAS has all functions listed on their web site with examples, but for a more accessible presentation of the functionality of SAS functions I strongly recommend *SAS Functions By Example* by Ron Cody. Furthermore, *Carpenter's Guide to Innovative SAS Techniques* by Art Carpenter is a comprehensive book which emphasizes how DATA step techniques can be used to solve complex data problems. Many of the examples in this book also cover topics relevant to the advanced certification which is discussed below.

One relatively large section is reading raw data and it is not as trivial as you might think – particularly if you are not doing it every day and getting all sorts of raw data problems to solve. The *SAS Certification Prep Guide Base Programming for SAS 9* is sufficient but if you need more examples it is recommended to have a look at *Reading External Data Files Using SAS: Examples Handbook* by Michele Burlew. This book prepares you for many different scenarios on how to read raw data into SAS.

The official certification prep guide SAS offers includes a small quiz at the end of each chapter, which is good to measure your initial understanding although the real exam questions are harder than those in the official guide, so just because you master the quizzes in the official guide does not necessary mean that the real exam will be as easy. SAS offers a practice exam for this certification but it does not come free, however it is very close to the real exam in terms of questions asked. It is recommended taking the practice exam shortly before the real exam, but with enough time to be able to go back to your books if you get an unpleasant surprise.

### **SAS CERTIFIED ADVANCED PROGRAMMER FOR SAS 9**

Although the techniques taught through the base certification are useful, the really interesting topics are covered by the advanced certification. This certification covers the use of SQL through the SQL procedure, the full macro language and advanced programming techniques.

The three main categories are comprehensive and require more preparation time compared to the SAS Certified Base Programmer for SAS 9 exam. The advanced programmer certificate actually tests your skills in three programming languages, the DATA step language, SQL, and the SAS macro language. Advanced programming techniques include topics such as creating indexes and when they should be used, combining data vertically using the SET statement, the INFILE statement and the APPEND procedure, hash objects as lookup tables, creating formats with the PICTURE statement in the FORMAT procedure. It also introduces you to efficient programming, how SAS uses memory, disk space and CPU time.

The advanced certification is indeed the more useful one but it is a requirement that the candidate must already have passed the SAS Certified Base Programmer for SAS 9 certification. For a more complete overview of the content the reader is encouraged to consult the SAS webpage ([sas.com](http://sas.com) webpage). At the time of writing the exam consisted of 60-65 multiple-choice and short-answer questions for which the candidate must score at least 65% to pass.

General one-fits-all advice is hard to give, considering each SAS programmer's different background, but there are several resources the candidate should consider: SAS has an official guide for the advanced programmer certification as well: *SAS Certification Prep Guide Advanced programming for SAS 9*. Though the book is comprehensive, it is recommended to take a look at the following books as well: *Carpenter's Complete Guide to the SAS Macro Language* by Art Carpenter is an excellent book on this topic, and SAS's own *SAS 9.4 Macro Language Reference* is not a bad choice either. *Carpenter's Guide to Innovative SAS Techniques* by Art Carpenter is also a book you should consider and this book is recommended as a preparation resource for the base programmer certification as well. The SQL procedure is well covered by the *SAS Certification Prep Guide Advanced programming for SAS 9* which makes it a sufficient resource to learn about this procedure.

The *SAS Certification Prep Guide Advanced programming for SAS 9* includes quizzes for each chapter as well, however it is recommended to purchase the practice exam offered here too as it gives a more realistic assessment of the candidate's skills.

### SAS CERTIFIED CLINICAL TRIALS PROGRAMMER USING SAS 9

This is the certification aimed directly at SAS programmers in the pharmaceutical and CRO industry by covering topics ranging from the clinical trials process, clinical trials data structures (CDISC standards; SDTM and ADaM), regulatory requirements through to exported data sets and concepts like last observation carried forward, basic statistical procedures like PROC TTEST and the validation of clinical trial data reporting. The SAS webpage offers a complete and detailed overview of the content (sas.com webpage). There is one aspect setting this certification apart from the others discussed here and that is that SAS offers two versions: One full version covering topics also in the base certification, and one accelerated version which contains fewer questions but which also requires the candidate to have passed the SAS Certified Base Programmer for SAS 9 exam beforehand. At the time of writing the exam consisted of 95-100 multiple-choice and short-answer questions for the full length version, and 70-75 for the accelerated version for which the candidate must score at least 70% to pass.

The preparation path for this certification looks a bit different compared to the base and advanced certifications. First of all, the topics should be selected to match your experience as a SAS programmer in the pharmaceutical industry, and secondly, SAS does not offer any official preparation guide or practice exam as for the base and advanced certifications but they do provide a comprehensive list of courses and material for preparation. The author has no experience of the courses offered by SAS and will therefore only discuss the literature: *SAS Programming in the Pharmaceutical Industry* by Jack Shostak should be on the shortlist of preparation material. In a very clear fashion the book walks the candidate through the fundamentals of working as a SAS programmer in the pharmaceutical industry. Many of the exam questions are covered by this book – in other words: highly recommended reading. Another book closely related to the above is *Validating Clinical Trial Data Reporting with SAS* by Carol Matthews and Brian Shilling. Validation is a big topic for a SAS programmer in the pharmaceutical industry, hence the value of this book. Many of the questions related to validation in the exam can be answered with the help of this book.

CDISC is an important part of a programmer's life in the pharmaceutical industry and as the use of CDISC develops to a standard in the industry, the importance of this knowledge can only increase. Candidates for this certification will find this part of the exam preparation – and the exam itself – easy or hard depending on their CDISC experience. Currently CDISC has one implementation guide for SDTM and one for ADaM (SDTM Implementation Guide v3.2 and ADaM Implementation Guide v1.1) both of which can be found at [www.cdisc.org](http://www.cdisc.org). For those with little or no experience, the guides might seem mysterious, but the concept of the standards is well described in the implementation guides.

As you prepare for this exam you have to line out your preparation based on your current knowledge and work experience – if you are a clinical database programmer you will probably not use the FREQ procedure to calculate Fisher's exact test or use the TTEST procedure to calculate statistics for paired samples very often. Fortunately *SAS Programming in the Pharmaceutical Industry* by Jack Shostak and other online resources are full of examples on PROC FREQ and PROC TTEST. And the same goes for candidates focused on ADaM and reporting who might need to spend more time on learning about the SDTM domains.

### SAS CERTIFIED STATISTICAL BUSINESS ANALYST USING SAS 9: REGRESSION AND MODELING

The newest of the four discussed certifications, the SAS Certified Statistical Business Analyst Using SAS 9: Regression and Modeling, was introduced in 2012 and as the (long) name suggests programmers working with clinical data is not the main audience. Nevertheless, it offers a structured path providing an introduction to statistics using SAS. For the experienced SDTM programmer with little or no daily exposure to the powerful statistical procedures in SAS, this certification will give that introduction. At the time of writing the exam consisted of 60 multiple-choice and short-answer questions for which the candidate must score at least 68% to pass.

The certificate covers topics like ANOVA, linear regression, logistic regression, the preparations required for predictive model performance and how to measure model performance. It will be clear whilst going through the content and preparations for this certification that some of the topics taught will be of little direct use for programmers occupied with SDTM, ADaM and table programming, but still the certification covers many general statistical concepts which are highly relevant for the statistical programmer working with clinical data. It should be said that the expectation is not that a programmer should be able to take on the role of an academically trained statistician, but to be able to produce more complex inferential statistics.

For the stat newbie it might be overwhelming to get started and to define a plan on how to study for the exam. Fortunately SAS provides a detailed overview on the content which should give the candidate a clear picture of what is expected (sas.com webpage). Assuming the candidate is relatively new to statistics and to SAS as a statistic tool it is strongly recommended to take the free course Statistics 1: Introduction to ANOVA, Regression, and Logistic Regression provided online at SAS<sup>1</sup>. It will not cover all of the different techniques and methods that might be in the exam but it does cover the basic assumptions behind many common parametric tests and more importantly it also includes many of the SAS procedures regarding syntax and functionality. What is not covered in the above discussed course is receiver operating characteristic curves, ROCs, which is going to be covered by at least 2-3 questions in the

---

<sup>1</sup> At the time of writing SAS offered this course for free (<https://support.sas.com/edu/schedules.html?ctry=us&id=2816>)

exam. It is recommended to have taken a look at Mithat Gönen's book *Analyzing Receiver Operating Characteristic Curves with SAS*.

### MOOCS

So-called Massive Open Online Courses offer plenty of exciting opportunities to learn and gain deeper understanding about different programming languages like R, Python, Java, subjects such as neuroscience, the basics of vital signs, how to design clinical trials, epidemiology and project management, to name but a few, all of which are all relevant to the programmer working with clinical data in the pharmaceutical industry or in a CRO setting.

This is not the place to give a detailed exposition, however, in 2016 MOOCs are an established structure for providing online education. On the general wikipedia.org page for MOOCs there are currently (August 2016) 23 providers listed (wikipedia.org webpage), consequently this paper will only be covering a fraction and MOOCs from coursera.org and edx.org will be discussed further below.

### BENEFITS FOR PROGRAMMERS IN PHARMACEUTICAL INDUSTRY

As has been briefly discussed above, the actual requirements for most programmer positions in the pharmaceutical industry who work with clinical study data go well beyond knowing the SAS programming language and having a decent understanding of CDISC standards, to take two examples. Often the programmer is exposed to complex study designs outlined in the protocol and statistical concepts for the endpoints in the statistical analysis plan. A programmer can of course learn about cross-over study designs, or bioequivalence clinical trials, or non-inferiority trials, the concept of randomization and much more from senior colleagues, well-developed learning management systems and the conventional web sources, however not all CROs and pharmaceutical companies are equipped with in-house developed courses provided through a learning management system that covers relevant topics, or senior colleagues who can spare a few hours to explain the concept and technical aspects of how to perform an analysis of survival data. Nor do all programmers have the time and the capacity to educate themselves in concepts largely new to them. This is the market gap MOOCs can fill. To give the reader a glance on what is offered, the author has selected five recently offered courses provided through coursera.org and through edx.org<sup>2</sup>.

Design and Interpretation of Clinical Trials, provided through Johns Hopkins University, covers the basics of randomized trials and how they are best reported. The course gives the student an introduction to terminology used in clinical trials and discusses some of the most widely used designs applied in clinical studies (coursera.org webpage). For the seasoned programmer with some 20 years of experience through the entire spectrum of clinical programming tasks this might all be very clear and obvious but for the less experienced programmer with only a few years of experience, a course like this can provide an excellent base to build knowledge and understanding of the process of clinical trials.

The understanding of different types of clinical trials and designs are one fundamental block, but understanding the statistics is as important. The course Statistical Inference, also provided by Johns Hopkins University through coursera.org, explains the rudiments of distributions, P-values and T-test powers and more (coursera.org webpage). Note that this is only one of many examples of basic courses covering many statistical concepts and which also provide hands-on exercises. Also, it should be mentioned that although they will not provide the in-depth understanding required to qualify as a statistician, these courses, offered through coursera.org and other MOOCs providers, will help the programmer to achieve a better understanding of the statistical concepts used in clinical trials.

The third and final example here is about programming. Courses using SAS software as the main programming tool are rare spices in the MOOC jungle, and for that we have the SAS certifications provided by SAS (as discussed above). True is that SAS might be the dominating tool to build databases from clinical trials data and to analyze the data, but there are still plenty of other programming languages used in the pharmaceutical industry, and knowing one or more of these can only be beneficial to the programmers' general understanding of coding. Programming for Everybody, University of Michigan, again provided through coursera.org, focuses on basic programming concepts using Python. How to code – Systematic Program design is a series of programming courses by The University of British Columbia provided through edX.org, another MOOC supplier. This is a hands-on course teaching the core of programming design and focuses on the general concept of programming rather than a specific language such as in the Python course above (edx.org webpage). The final example here is again from Johns Hopkins University and is a pure R language programming course. The course walks the student through the practical aspects of statistical computing in R, covering fundamentals as reading data into R, accessing different R packages, debugging and much more (coursera.org webpage).

### GET THE MOST OUT OF YOUR MOOCS

For many but not all MOOCs it is possible to pass all quizzes and assignments with a minimum of effort but your return on the invested time might still be very poor. In order to get the most out of your invested time and effort a plan is advisable. First of all it is highly recommended to understand what it is you want to learn and what you can learn from a course delivered in this format with a limited time frame. This will not only help you structure your learning but will also help you to manage your expectations. Secondly, it is important to be active. It might be tempting to watch one or two video lectures after work to make progress with the course and to reduce your guilty conscience for not

---

<sup>2</sup> This is not an endorsement of coursera.org or edx.org but merely mentioned for illustrative purposes.

## PhUSE 2016

doing enough. With that strategy it is likely that you will go through the course without actually learning anything, or you might learn a few new things you could have easily learned by paying extra attention while reading the Wikipedia article on the same topic. Instead, be active, take notes, pause the video lecture to think, or if you are taking a practical course, like a programming course, pause and switch to your editor and type in the code just shown to you.

### WORD OF CAUTION

Undoubtedly, the MOOCs provide the programmer in the pharmaceutical industry, or many other professionals, a broad range of high quality educational material and easy access to new knowledge. However, based on the author's own experience, it is very easy to get carried away and sign up to multiple courses simultaneously, which can be justified but with significant obligations as full time work, family and more there is a high risk that, the weekly series of 10-15 minutes lectures, quizzes and assignments get squeezed in to an already full schedule and that you eventually drop out from the courses.

### HANDS-ON EXPERIENCE

Whilst the SAS certifications will provide a solid foundation of SAS programming skills, and MOOCs can be a worthwhile stepping stone to in-depth understanding of a wide range of technical topics useful in this profession, many skills can only, and are best, taught by actually doing them in real life. Acquiring new skills can also be a painful exercise but the programmer must leave their comfort zone, and challenge not only themselves but also the people around them to appreciate this new opportunity of strengthening the team.

There are a few things to be aware of in order to get the most out of the time invested. Firstly it is crucial that you define what it is you would like to achieve. Becoming the ADaM expert of the entire department in less than 3 months might not be possible, or advancing to an in-house SAS macro developer with only 6 months of SAS programming experience might also be too ambitious.

It is crucial to match your current level with what you can actually do. It depends much on the company or the department so it is important that whatever you set out to accomplish it can actually be realized.

While some skills can be developed through hands-on experience as they are integrated into your daily work tasks, like for example becoming a specialist in using PROC SQL, most of them require you to involve co-workers and most importantly your manager. It is your manager who can support you and help you to find the right tasks or projects for your goal. Assuming the goal is to advance from being a support programmer to a lead programmer who is the one assigning tasks to the programming team, approving vacations, reviewing the specifications and tables, communicating with the statistician, the medical writers and the client, then it is absolutely crucial to gain the support of your manager. In order to better understand obstacles and opportunities let's look at two fictional, but highly relevant situations.

### THE MACRO DEVELOPER

Imagine a fairly experienced programmer specialized in SDTM programming but a novice in the SAS macro language who wants to transfer into a macro developer role in the company. The first step is to identify the goal, and define what exactly can be achieved with the current circumstances. Each company, CRO or sponsor, depending on size and other circumstances offers different opportunities and paths. So, it is important to find out if the company has a specialized team or sub-department in the statistics/programming department working full time to develop and support the company's SAS macro library. Or are macros developed and maintained by functional sub-teams, for example an SDTM or an ADaM team. It is important to identify the possible paths available in your company or department, that way you can more easily sell it to your manager and gain the support required to achieve the goal.

The second step for the aspiring macro developer is to introduce the goal to the manager and explain why and how this undertaking can be realized. It is of course important for the programmer to stress the advantages for the department of having another macro developer in the team and also what sets you apart, depending on previous experience, compared to the other macro developers (or potential macro developers). Nevertheless, not all goals for professional development can be met at once and it might be that there is no need, demand or budget for a new macro developer, however, it does not mean that the door to this path is closed forever; you might settle for a 50% position as a macro developer for the first months, or that the case is reviewed again in 6 months. At least you've now got your manager's attention.

### THE LEAD PROGRAMMER

The second example is a bit different as it covers a transition from a support programmer into a lead programmer role. Depending on the company's size, CRO or sponsor, the paths and opportunities to advance from a good support programmer to a lead programmer may vary significantly. It is fair to assume that in a large CRO the volume of clinical studies is much greater, so the demand for lead programmers is going to be higher compared to a small CRO or a sponsor where much of the work is outsourced and the programmer's daily tasks might be different from the programmer working for a CRO. For the example below we assume a CRO setting.

The first thing to do is to identify the tasks falling in the domain of the lead programmer, which might consist of: being a technical mentor for junior programmers, writing and reviewing others' specifications for SDTM or ADaM domains, reviewing the SAP and communicating with the client. The lead programmer's tasks are in no way limited to the items listed above but can be some of the key responsibilities for this role. One of the significant differences between

## PhUSE 2016

support and lead programmers, worth stressing here, is the amount of communication. While the support programmer offers mainly technical skills in the form of programming SDTM and ADaM data sets, tables and figures, the lead programmer must have the skills needed to lead a small team of programmers and communicate with different functions in the own company and with the client.

The next action, assuming the tasks assigned to the lead programmer are still appealing, is to make manager aware of the goal. In some companies, particularly CROs, it should be easier to take this step compared to transferring into an in-house SAS macro developer position, but still it can be difficult to find the perfect project immediately. Initially, it might also be possible to transfer into a sub-lead position for larger projects in order to collect the necessary experience.

It should be clear that the opportunities are not limited to the two examples above but can practically include any type of role in your department or company.

### CONCLUSION

SAS certifications, MOOCs and hands-on experience are not in any way the only paths leading a programmer to new skills and new levels. There are many different ways that knowledge and new skills can be taught. And one particular approach does not exclude another- it is an excellent strategy to combine certain well selected MOOCs while exploring new tasks in your daily work, or to combine learning for a new SAS certification and simultaneously make use of the new knowledge at work.

### RECOMMENDED READING

ADaM Implementation Guide v1.1. [www.cdisc.org](http://www.cdisc.org)  
Burlew, Michele. 2002. Reading External Data Files Using SAS: Examples Handbook. SAS Institute  
Carpenter, Art. 2004. Carpenter's Complete Guide to the SAS Macro Language, 2nd ed. SAS Institute  
Carpenter, Art. 2012. Carpenter's Guide to Innovative SAS Techniques. SAS Institute  
Cody, Rod. 2010. SAS Functions By Example, 2nd ed. SAS Institute  
Chris Holland & Jack Shostak. 2012. Implementing CDISC Using SAS: An End-to-End Guide. SAS Institute  
Carol Matthews and Brian Shilling. 2008. Validating Clinical Trial Data Reporting with SAS. SAS Institute  
Gönen, Mithat. 2007. Analyzing Receiver Operating Characteristic Curves with SAS. SAS Institute  
Shostak, Jack. 2005. SAS Programming in the Pharmaceutical Industry. SAS Institute  
SAS Certification Prep Guide Base Programming for SAS 9, 3ed. 2011. SAS Institute  
SAS Certification Prep Guide Advanced programming for SAS 9, 3ed. 2011. SAS Institute  
SDTM Implementation Guide v3.2. [www.cdisc.org](http://www.cdisc.org)Contact

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

Hannes Engberg Raeder, Principal Statistical Programmer  
PRA Health Sciences  
Gottlieb-Daimler Strasse 10  
68165 Mannheim, Germany  
[engbergraederhannes@prahs.com](mailto:engbergraederhannes@prahs.com)  
[prahs.com](http://prahs.com)

Brand and product names are trademarks of their respective companies.