Be a Dead Cert for a SAS® Cert

How to prepare for the most important SAS Certifications in the Pharmaceutical Industry

Hannes Engberg Raeder, inVentiv Health Clinical, Germany

ABSTRACT

It is a fact that SAS has become the standard for analyzing and reporting clinical data in the pharmaceutical industry, thus it is one of the most important tools for the clinical database programmer, statistical programmer and statistician. Getting certified raises your overall programming awareness, improves the quality of your work and, earns official recognition from SAS. There are currently three certifications of relevance for the clinical SAS programmer; SAS Certified Base Programmer, SAS Certified Advanced Programmer and SAS Certified Clinical Trials Programmer. However, lack of time and good learning strategies are potential obstacles that may discourage candidates from trying. Based on my own experience, the aim of this paper is to provide an overview, resources for self-study and advice on dealing with practical aspects like integrating learning into daily work and on the exams themselves such as the type of questions encountered and exam strategy.

INTRODUCTION

On the SAS web site you can read that *SAS is the de facto industry standard for clinical data analysis and reporting in the life sciences/pharmaceutical industry*, and not surprisingly SAS is de facto the most important tool you have as a clinical database programmer, statistical programmer or statistician in the pharmaceutical industry[1]. This paper will focus on, what I believe to be, the most important SAS certificates for SAS programmers in the pharmaceutical industry, namely:

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

Learning and preparing for a SAS certification is a rewarding experience where you will

- be exposed to new techniques, procedures, statements and options
- raise your awareness of your own programming
- increase the quality and productivity of your work and expand your SAS programming toolbox
- earn official recognition from SAS for your knowledge

Nevertheless, most SAS programmers are working full time which make it difficult to find time for learning and preparation. But you do not need to spend every evening and weekend for six months preparing, and it does not need to be expensive – with four to five hours a week and some well-chosen SAS publications these certifications are manageable for all SAS programmers – beginners and seasoned ones alike.

This paper is based on my own experience learning and preparing for these certifications and consists of three parts. The first focuses on the certifications themselves – a brief overview of the content as well as resources for self-study. The second part focuses on how to prepare and on the practical aspects of learning. The third part focuses on the exam and situations to be aware of at the exam such as allocation of time and types of questions to expect.

---

[1] For simplicity the term SAS programmer will be used from here on to denote any role that includes use of SAS products to handle, prepare and analyze data from clinical trials.

PART ONE: THE CERTIFICATIONS

SAS Certified Base Programmer for SAS 9 and SAS Certified Advanced Programmer for SAS 9 are two certifications that will greatly strengthen your SAS programming toolkit and raise your overall programming awareness[2]. For the SAS programmer working with clinical data – preparing, analyzing and reporting – these are the cornerstones of your programming toolbox.

Although you can pass the base programmer certification and then the accelerated version of the clinical trials programmer certification, skipping the advanced programming certification, strong DATA step and macro language knowledge, which are key topics in the advanced, are required for the clinical trials certification. Hence, I highly recommend you to take both the base and the advanced programmer certifications before taking the clinical trials programmer credential.

SAS CERTIFIED BASE PROGRAMMER

This exam has 64 multiple-choice and short-answer questions. To pass the exam you must achieve a score of 70% in 110 minutes. The main focus is the DATA step, for example when to use the SELECT statement instead of IF-THEN/ELSE statements, how to use arrays, read in different structures of raw data, convert from numeric to character and, apply formats and informats, and so on. Some of the topics covered are:

- Accessing data, combining datasets, access an Excel workbook
- Creating data structures
- Managing data, modifying variable attributes, process data using loops and arrays
- Generating reports, ODS
- Handling errors, resolve logical, syntax and data problems

For a detailed syllabus see SAS Institute's web site www.sas.com/certify/creds/.

WHAT TO STUDY?

SAS offers an official guide called *SAS Certification Prep Guide Base Programming for SAS 9* which covers the subjects relevant to the exam and each chapter includes a small quiz in the end. The guide is comprehensive and detailed, although the real exam questions are in my opinion harder than to those in the official guide. Depending on your current experience and work tasks involving SAS 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 structured presentation of the functionality of SAS functions I strongly recommend *SAS Functions By Example* by Rod Cody.
- *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. You will find many questions from both the base and advanced certification covered in this book.
- Reading raw data is not as trivial as you might think – particularly if you are not doing it every day and geting all sorts of raw data problems to solve. In my opinion the *SAS Certification Prep Guide Base Programming for SAS 9* is sufficient. Though, if you need more examples I recommend *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.

---

[2] This paper only discusses SAS Version 9. The version number will not be written out when discussing the certifications.

SAS CERTIFIED ADVANCED PROGRAMMER

The foundations of the advanced certificate are the SQL procedure, the SAS macro language and advanced SAS programming techniques. The exam contains 63 multiple-choice questions and you must score at least 65% in two hours to pass. In order to receive the advanced programmer credential, you must already have gained the base programmer certification. Areas covered are:

- Accessing data using SQL
- Macro processing
- Advanced programming techniques

These three categories are comprehensive and require more preparation time compared to the base programmer certificate. The advanced programmer certificate actually tests your skills in three programming languages, the DATA step language, SQL which is introduced here, 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. For a detailed list of requirements see SAS's web site.

The SAS Certified Advanced Programmer certification is comprehensive and therefore a highly rewarding experience.

WHAT TO STUDY?

General one-fits-all advice is hard to give, considering each SAS programmer's different background, but there are several resources to 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, I recommend taking a look at the following books as well:

- To boost your SAS macro language knowledge: *Carpenter's Complete Guide to the SAS Macro Language* by Art Carpenter is an excellent book on this topic, and SAS's own *SAS 9.3 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. I recommend this book as a preparation resource for the base programmer certification as well.

In my opinion, the SQL procedure is well covered by *SAS Certification Prep Guide Advanced programming for SAS 9,* therefore I have chosen not to recommend alternative literature even if there are comprehensive options available.

SAS CERTIFIED CLINICAL TRIALS PROGRAMMER

This is the newest of the three certifications discussed in this paper and is aimed directly at SAS programmers in the pharmaceutical industry. Currently there are two versions offered: One, accelerated, version for those with the base programmer credentials and one for those without. The accelerated version contains 71 questions and you have to pass 70% of them in two hours. The full length clinical trials programmer certification has 99 questions and 70% must be achieved in three hours to pass and you do not need to have the base programmer credential.

I will focus on the accelerated version since I assume that most SAS programmers will take the base certification before going for the clinical trials programmer certification. The accelerated version covers the following topics:

- Clinical trials process, SAP, principles of 21 CFR Part 11, Good Clinical Practices
- Clinical trials data structures, CDISC

- Import and export clinical trials data
- Manage clinical trials data
- Transform clinical trials data, change from baseline, LOCF
- Apply statistical procedures for clinical trials
- Macro programming for clinical trials
- Report clinical trials results, PROC REPORT and ODS
- Validate clinical trial data reporting

As much as a SAS programmer in the pharmaceutical industry needs to know DATA step, PROC SQL, the SAS macro language and at least a couple of useful PROCs, understanding of the data, the regulations, documentation and, a basic understanding of statistics is just as important.

WHAT TO STUDY?

The preparation path for this certification looks a bit different compared to the base and advanced certifications. First of all, the topics are 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. I have 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 your list of preparation material. In a very clear fashion the book walks you 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 SAS 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 v3.1.3 and ADaM Implentation Guide v1.0) both of which can be found at 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.

PART TWO: HOW TO LEARN AND PREPARE

During my own preparation for the base programmer certification I did not have a plan at all – except just to read and try to memorize everything. The lack of a plan made my preparation very cumbersome and I could have been much more efficient. However, at the end of my preparation I deliberately started to implement all of the things I learned. I became more aware of my programming and tried to make my programs shorter, faster and less likely to contain any logical errors.

As I started to learn and prepare for the advanced certification I was keen not to spend that much time and energy again. I managed to pass both the advanced and clinical trials programmer certification in less than four months – scoring 86% and 97% respectively, spending an average of five hours a week on self-study outside work time during this period by implementing the strategies described below.

ASSESS YOUR CURRENT KNOWLEDGE – OR LACK THEREOF

It is critical that you figure out what you know – and what you do not know. Make an inventory of your current knowledge, related to the certificate you are preparing for, and categorize it into knowledge of techniques and methods you know well, techniques and methods you know of and you can use if you consult manuals, books or other sources and, techniques and methods you do not know at all.

Assume a candidate is preparing for the SAS Advanced Programmer certification and is a passionate DATA step programmer who knows all the related statements, options and the most common procedures like a book this clearly go to the category techniques and methods this programmer know well. It is more difficult with techniques and methods you can use with the help of a book or example programs. Every SAS programmer could fill pages with functions, macro functions, formats, informats, options and procedures they know of and occasionally use after consulting an old program, a manual, a paper or a colleague but without aid the SAS programmer might fail to answer a question on that subject. As an example, if you need to debug a macro you know there are several options you can use, like the MPRINT and the MLOGIC system options. If you use one or both of them you may well find your bug and resolve the problem. However, if you are only shown a small part of a log and asked which option produced it, it might be tricky to deliver the correct answer.

Make absolutely sure you correctly assess your current knowledge on the exam you are about to take otherwise you are at risk of having this information delivered to you on the exam day by not passing.

LEARN 8 HOURS A DAY OR MORE – HOW TO INTEGRATE LEARNING INTO YOUR DAILY WORK AS A SAS PROGRAMMER

As a full time working SAS programmer it is possible to learn 8 hours per day – by integrating the new techniques and methods you learn into your daily work. By doing so you will not only learn and master new techniques but also retain what you learn. The idea with the certifications is not that you just memorize for the exam and continue to use the old techniques you are used to after passing it, but continually build up your SAS programming toolkit with the new tools and techniques you encounter, and make use of them.

If you prefer to merge data sets using the MERGE statement, you should start using PROC SQL to join your tables instead – necessary for the advanced certification. If you need proof that merging with DATA step produces the same result as an SQL join given that they are used correctly; use the COMPARE procedure to compare the output datasets from both techniques and learn to interpret and modify the PROC COMPARE output at the same time. Questions about PROC COMPARE output are likely to be included in the clinical trials programmer certification exam.

For the SQL geek it is time to take a look at some of the more advanced DATA step techniques. The *SAS Certification Prep Guide Advanced programming for SAS 9* includes examples of SET statement with the KEY= option, the END= option, the MODIFY and UPDATE statements, the use of a DO UNTIL loop to generate

summary statistics, and hash objects. Furthermore, if you always sort your data sets using the SORT procedure, start using PROC DATASETS or the DATA statement to create indexes. This might not always be the most efficient way to program, but during the exam, in this case the advanced certification, you need to know how to do this. There is no better way to learn than applying the new techniques and methods in your daily work.

GET USED TO READING CODE – ACCURATE AND FAST

Many of the questions in the exams have code examples. Sometimes short fragments and sometimes lengthy SQL queries or DATA step code. Compared to your default SAS editor the code examples in the exam are not color coded, but just plain black text on a grey background. Display 1 below illustrates the difference between the visual appearance of the code in the exam and in the default enhanced SAS editor. The code in display 1 is just an example but similar to questions asked in the real exam (in this case the advanced programmer certification). In the example question below you are asked to select one of four macro calls to the dsprint macro and only one of them is correct. This requires you to carefully read and understand the code in order to answer the question correctly. For the SAS programmer used to working with the color coded editor it can take longer to understand the code and the risk of overlooking a syntax error might increases.

```
%macro dsprint(ds,var);
  %if &var= %then %do;
    proc print data=&ds;
      title "Full Listing of %upcase(&ds) data set";
    run;
  %end;
  %else %do;
    proc print data=&ds;
      var &var;
      title "Listing of %upcase(&ds) data set";
    run;
  %end;
%mend dsprint;
```

```
%macro dsprint(ds,var);
  %if &var= %then %do;
    proc print data=&ds;
      title "Full Listing of %upcase(&ds) data set";
    run;
  %end;
  %else %do;
    proc print data=&ds;
      var &var;
      title "Listing of %upcase(&ds) data set";
    run;
  %end;
%mend dsprint;
```

Display 1. Example of the difference between exam layout (left) of the questions and the default enhanced editor.

The best and most efficient way to get used to reading code as displayed to the left is to do some coding without any visual color aid in the editor. This will strengthen your ability to recognize missing semicolons, misspelled words, and invalid statements without relying on an editor to highlight those for you.

Another way to speed up your code reading skills is to perform programming tasks by writing the code solution on paper. Make sure you base your task on some real data and when you are done you can check it by running the code in SAS. This will train you to rely less on the context you are used to and to spot syntax errors much faster.

BE FOCUSED – FOLLOW A SCHEDULE

Create a plan for your exam preparations. It is important to be practical – you will not be exam-ready in 10 days but you will not need 10 months either. Your current tasks and experience are factors that impact how much time you will need. You might need to study more in your spare time if you cannot integrate the learning into your daily work. If this is the case I recommend that you make a detailed plan for those topics and remember to keep your new knowledge alive. Unfortunately, if you just read and never practice what you read, it will soon be forgotten.

Be detailed in your plans – it makes it harder for you to deviate and helps you to keep an overview as you progress. Concerning the content of your learning plan it should reflect your SAS knowledge inventory (discussed above) – in that way you will maximize the outcome of time spent reading and practicing.

When drafting your plan you should be realistic – to read 40 pages and actually retain even some of this knowledge is often hard after a day's work. It is always better to strive for quality and not quantity in your learning and preparation phase. Being over ambitious will only harm your motivation in the long run.

THE PRACTICE EXAMS

According to SAS the practice exams test the same knowledge as the real certification exams. Since the format of the questions in the practice exams is the same and all topics are proportionally weighted to match the real exams the practice exams work well as predictors for the outcome of the real exam. The exams include 50 questions.

One disadvantage of the practice exams is that you can only purchase one set of questions. If you score 90% after doing the practice exam eight times the prediction of the outcome from the real exam becomes questionable. My advice is to prepare and learn by following your study plan (discussed above) and the official guides (*SAS Certification Prep Guide Base Programming for SAS 9* and *SAS Certification Prep Guide Advanced programming for SAS 9*) and only just before you schedule the certification exam you take the practice exam. This will boost your confidence as you see that your effort is paying off, give you a valid forecast of how you would have performed if it would have been the real exam, and make you aware of any topics you need to brush up on.

In order to get the most out of the practice exam after finishing it, you should take your time and go through the questions and your answers. Say you gave the wrong answer for two macro questions, and looking carefully at them it might be that you did not fully understand the concept of the local and global macro variable tables, or you were wrong on one or two DATA step questions because you did not pay attention when SAS automatically assigned attributes to variables. Furthermore, make sure that you know why you answered a certain question correctly. If you have to guess between two alternatives which both seemed likely you should review that topic again.

PART THREE: AT THE EXAM

After having prepared successfully for one of these certifications and scored well in the practice exam you are now ready for the real certification exam[3]. It is probably fair to say that the outcome of the exam in most cases is determined before you enter the test facility thanks to your experience as SAS programmer and your preparation. But, there are a couple of simple 'rules' to keep in mind during the exam that can influence your result for the better.

STRATEGY FOR ANSWERING QUESTIONS

Do not get stuck on one question by reading and trying until you believe you have the correct answer. This can happen with questions looking like the example in display 2 where you are supposed to pick the answer that best describes the content of the output data sets. To be able to answer the question you have to 'iterate' manually through the data taking the logic in DATA step into account and sum the number of observations written to each of the output data sets in the DATA statement.

---

[3] For information about where and when you can take the exam, fees and re-take policy I advise you to consult SAS's web site.

```
SUBJIDN   DBP    SBP

1000114   93     157
1000116   82     144
1000117   92     140
1000120   71     116
1000128   87     146



data norm high check;
  set bp;
  if sbp gt 135 and dbp gt 82 then output high;
  if sbp lt 118 and dbp lt 80 then output norm;
  else output check;
run;


What is the result?
```

Display 2. Example of a question that needs careful attention

If you encounter a question you need to spend more time on and the answer is not coming to you, leave that question for later and move on. The reason is simple; you might spend valuable time trying to reach an answer which is not correct anyway and use up time you could have used for other questions, and even if you temporarily leave the question there is a good chance that the answer will come to you after some time or some other question will trigger your memory. In the meantime there are questions like the one below that, given that you know the answer, can be solved after just a couple of seconds.

```
Which function calculates the average of the variables Var1, Var2, Var3, and Var4?

A) mean(var1,var4)
B) mean(var1-var4)
C) mean(of var1,var4)
D) mean(of var1-var4)
```

Display 3. Example of a question that can be answered quickly

If you know the syntax of the mean function you will immediately know the answer to the question in display 3. You do not need to take data into consideration or read lengthy SQL clauses or DATA step code. On these questions you can save a lot of time. Consider the base certification exam which includes 64 questions to be finished in 110 minutes. That gives you in average 1.71 minutes to spend on each question – it might sound scary but it is more than enough, as many questions are fairly quick fire.

If you do not know an answer even if you have tried long and hard, do not worry. Most of the questions are multiple-choice with four alternatives where only one can be correct. If you can identify one or two of them as wrong you have considerably raised the probability of answering correctly if you need to guess.

SHORT-ANSWER QUESTIONS

The exams do not only contain multiple-choice questions, but also short-answer questions. The candidate has to provide the answer. A typical question might look like in the example in display 4.

```
Complete the following INPUT statement by using the line-hold specifier that
holds the current record for multiple iterations of the DATA step.

data WORK.june90;
    infile temp;
    input Date : date. Htemp ?
```

Display 4. Example of short-answer question

In this example you should complete the INPUT statement with line-hold specifier that holds the current record. The short-answer questions are no harder than multiple-choice questions given that you know the answer. The

multiple-choice questions might even confuse more if two or more alternative appear to be correct. However, make sure you only put what is asked for in the short-answer field without extra spaces or semicolons. You do not want to risk a correct answer not being counted.

THE MARK FUNCTION

During the exam you can mark any question and later on once you have gone through them you will get an overview of questions you have marked and you can go back and review them again given that there is still time left. I strongly encourage you to use the mark function but make sure you do not overuse it by marking too many questions. If you have 60 out of 70 questions marked in the end it will probably be of little use. However, used with moderation this functionality is very helpful.

THE RESULT

The exam result will appear on the screen as soon as you finish the exam (or run out of time). You will be shown your percentage of correct answers and a bar displaying your percentage in relation to the pass mark. You will also see the percentage for each topic.

CONCLUSION

SAS programmers in the pharmaceutical industry, beginners and seasoned ones alike will benefit from learning and taking these three certifications. The base programming certification is the stepping stone to the advanced programming certification and the clinical trials programmer certification.

The efforts needed for learning and preparation are highly individual and each candidate should undertake a detailed inventory of her or his own SAS skills at the outset. The outcome of that inventory should form the basis for the study plan. The plan will give you an overview and help you to organize your learning and preparation. It is recommended to purchase the practice exams available for the base and advanced programming certification.

To get certified is an excellent way not only to grow the set of tools you use as a SAS programmer in your daily work, but also to earn official recognition from SAS.

REFERENCE

ADaM Implentation Guide v1.0. 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
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 v3.1.3. www.cdisc.org
http://support.sas.com/certify/

CONTACT INFORMATION

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

Hannes Engberg Raeder, M.Sc.
inVentiv Health Clinical
Berlin, Germany
hannes.engbergraeder@inventivhealth.com or hannesengberg@gmail.com