PharmaSUG2011 - Paper IB01

Success As a Pharmaceutical Statistical Programmer

Sandra Minjoe, Octagon Research
Mario Widel, Roche Molecular Systems

ABSTRACT

Why do some people excel in their careers? This is a difficult question to answer and one that is especially puzzling when you are just starting out or aren't doing as well as you expected. The authors of this paper have been in the industry for a combination of more than 30 years, and have observed many successful and not-so-successful pharmaceutical statistical programmers. This paper is a compilation of our experiences and expresses our opinions of the skills needed and best suited for the role of pharmaceutical statistical programmer. It can help you identify your skill gaps and describe what you can do about them.

INTRODUCTION

As with any career, there is a set of specific skills that is useful for the statistical programmer. Some of these skills can be acquired, such as by taking courses in school, reading books, and researching on the internet. Other skills are more inherent, and relate back to what drives or excites us. And finally, there are analytical skills, where we apply what we've learned and get work done through others. For each of these three categories, acquired, inherent, analytical, we outline some specific skills that are useful to the pharmaceutical statistical programmer, explain how the skill is used in the role, describe what it looks like when there is a skill gap, and offer some suggestions on where to learn more about the skill.

DEFINITION

When we use the term "successful" in this paper, we are not implying a large salary or even a lofty job title. Rather, we mean programmers who are well-respected, who others want to work with, and who demonstrate a sense of excitement and satisfaction in what they do.

ACQUIRED SKILLS

Just like a baker needs to know how to bake, and a singer needs to know how to sing, a programmer needs to know how to program. And like those other professions, there are various sub-categories, such as a pastry chef or opera singer. The same goes for statistical programming in the pharmaceutical industry. We need the generalist skill of programming, plus more specific skills in the sub-categories of pharmaceuticals and statistics.

We can acquire or learn many skills in various ways. Just as a successful baker or singer learns their profession with a combination of rote work and many years of practice, a statistical programmer can acquire our skills in different ways.

Programming Skills

Pharmaceutical companies and CROs typically use SAS® to analyze clinical trials data, and statistical programmers often spend much more time preparing datasets for analysis than performing actual statistical analyses on them. Thus a good working knowledge of Base SAS®, especially in the Data Step®, is essential.

Interestingly, statistical reviewers at FDA, who are often the customers of our work, usually use other packages, such as JMP®, S-Plus®, R®, i-Review®, to do their work. Because FDA reviewers aren't programmers, they are often not as comfortable as we programmers with data manipulation. We must prepare datasets in such a way that even a non-programmer can use them to perform analysis. Our goal is usually to make our datasets "analysis ready", meaning that the data doesn't need to be transposed, merged, or otherwise transformed before use in a statistical procedure. This is why we need to do so much work before ever getting to the actual statistical analysis.

As for performing the statistical analysis, we also need to know various components of SAS/Stat®. Because the types of statistical analyses vary, it can be hard to prioritize these specific statistical procedures. In some areas, survival analysis is common, in others regression or linear analysis is more typical. Thus an ability to learn and apply statistical procedures is at least as helpful as knowledge of any specific subset of statistical procedures.

Other important SAS knowledge includes macros and reporting techniques. Macros are important because they allow us to shorten and standardize code. Some companies use output procedures such as PROC REPORT and PROC TABULATE, while others prefer the flexibility of DATA _NULL_ for producing pretty output.

SAS skills can be learned in a number of ways. Some of us programmers take training classes to learn different components of SAS, such as offered by SAS Institute[1], at local colleges and universities[2], or through other private companies[2]. Others have picked up SAS as a by-product of courses taken in a degree program. And some have learned on their own, either on-the-job or by using manuals and other programming books. All are viable options, so we can choose the method that works best for us based on our learning style and finances.

Statistical Skills

It might seem that a degree in statistics would be required for the position of statistical programmer, but this is usually not the case. While most successful statistical programmers have had some college-level statistics, often it is only a class or two. The authors have met successful statistical programmers who have degrees in nursing, psychology, physics, journalism, and even forestry. Regardless of the degree, the story is often the same – in their degree program they learned SAS while taken a statistics class or two required by their major or minor, enjoyed it, and after graduation ended up finding a job in that area.

At a minimum, a statistical programmer needs to understand descriptive and basic inferential statistics, because these are used regularly. Survival analysis, ANOVA and regression analysis are also common in the industry, and a general understanding of these concepts is beneficial. More and more, Bayesian analysis is being incorporated. A statistical background helps us to research and determine the most appropriate procedure for the required analysis.

Those who end up in the field with little or no statistics background will find it helpful to take a course or two in statistics to become more comfortable with the terminology. Local colleges and universities often offer basic statistics and biostatistics courses, often in the evenings, which is convenient for working adults.

Clinical Skills

Many statistical programmers will argue that we don't need any skills in the clinical area. These authors believe that this attitude limits success. Instead, we recommend that the statistical programmer make an effort to learn as much as possible about the clinical side.

Some questions you might investigate include:

- What is the disease, and how are patients affected by it?
- What is the medical need? Is the goal to make a drug to cure the disease, prolong life, cause fewer side effects, or just to make the patient's life more livable?
- How does the treatment under study work? What is the mechanism of action?

A lot of this information can be found within the company itself. For example, a study protocol typically describes the disease, medical need, and mechanism of action. However, this text is often written by and for medical professionals, and can be a bit complicated for the lay person. The company's annual report often contains a lot of this same type of information, and because it is written for investors it can be much easier for us to understand. If the company has a marketing department that produces literature for patients, that can be a great resource. In this age, the internet is a wealth of information, and we can learn a lot about the disease and medical need from the company's website, WebMD®, Wikipedia® or even Google®.

It's also helpful for the statistical programmer to know some chemistry for lab unit conversions. A lot of laboratory data is collected in clinical studies, and we are often required to do some rather tricky transformations in order to produce the results needed for analysis. Many unit conversions, such as from mg to g, are straightforward, and only require basic math and programming knowledge. Others are more complicated. When converting to or from moles, does it surprise you that the conversion factor varies depending on the substance involved? Do you know, when the unit is percent, what is it a percent of? If you work with a lot of lab data and are having trouble understanding these concepts, consider taking a chemistry course.

Industry

Successful statistical programmers are in touch with the industry and how it is evolving. They are not surprised by new guidance documents issued by FDA[3] or CDISC[4], but have been anxiously waiting for them to be posted. They attend conferences and meetings with others in their profession, read publications written by others, often give presentations or write papers, and get involved in industry groups.

The Clinical Data Interchange Standards Consortium[4] (CDISC) has several teams and sub-teams that have been developing standards for the pharmaceutical industry, and these teams are made up of those of us who work in the industry plus also volunteer our time on this work. CDISC standards are becoming more and more accepted within the industry and by FDA. Most pharmaceutical companies across the world have at least begun to discuss how and when to convert their data into CDISC structures, and many now use CDISC as an internal standard. All statistical programmers in the industry should be familiar with the ADaM and SDTM models, and the documents produced by these sub-teams are available for download. CDISC holds interchanges in different areas of the world, where

attendees can present and attend papers, network with other programmers and FDA reviewers, and learn about the current and upcoming CDISC standards. Some areas also have local CDISC groups[5] that meet throughout the year. Training on these standards can be taken directly through CDISC or from other vendors.

The Drug Information Association[6] (DIA) covers a wide range of topics in the industry. They have many regular meetings, including some specifically related to programming. For example, each spring there are annual meetings for Electronic Document Management, Computational Sciences, and Statistics. DIA also offers training courses on topic areas including data management, electronic submissions, and project management. They also produce many publications, including a free daily briefing e-newsletter that summarizes the current events of industry companies.

PharmaSUG[7] and PhUSE[8] are focused specifically on statistical programming in the pharmaceutical industry. Both offer regular meetings where we can attend/present papers and posters and network with others in the industry. PharmaSUG has thus far held conferences only in the United States and up until recently PhUSE's meetings and conferences were focused in Europe. PhUSE maintains a Wiki, where industry programmers can read about and add to topics, such as how to impute partial dates and avoiding hard-coding. PhUSE also produces a peer-reviewed journal for members.

For more general SAS knowledge, there are other SAS users groups[9] a programmer might consider attending. SAS Global Forum, the largest of the general users groups, has a spring conference and always includes papers specifically related to the pharmaceutical and health science area. There are regional SAS users groups spread across the country, and these groups each typically hold an annual meeting in the fall. There are also local SAS users groups that hold meetings throughout the year.

For more general statistical knowledge, consider the American Statistical Association[10] (ASA). They have an annual Joint Statistical Meeting (JSM), plus many local chapters offer other meetings. Like other groups mentioned above, this gives industry statistical programmers a chance to attend and present papers and also network with each other. ASA provides some career and education information and produces peer-reviewed journals and other publications. Those with statistical knowledge can volunteer within ASA, such as with their K-12 statistic education program.

There are many opportunities to get involved in industry groups. More experienced statistical programmers should consider contributing to standards and the wealth of publication opportunities. Even those new to the industry can attend meetings, listen to presentations, read papers and other publications, and network to learn more about what is happening in our industry. With a little initiative, statistical programmers can find or even create an industry group to suit their needs.

Advanced Degree

Although there are exceptions, the authors have noticed that the more successful statistical programmers seem to have a master-level degree. As mentioned earlier, many successful statistical programmers actually learned SAS while in school, specifically in a statistics course that was a requirement for their masters-level degree. The authors concede that there might be a large confounding factor concerning this type of education, as many companies require and/or reward for advanced degrees.

For those who do not have an advanced degree, it might not be required to get one to be successful. Keep in mind, however, that it can be more of a struggle to convince others of our value without one. There are now options for working adults, including evening and online degree programs such as offered at Texas A&M[11].

INHERENT SKILLS

Unlike acquired skills, which by definition are skills we can learn, some skills are more inherent. For example, no matter how much musical training you receive, you must also have some inherent skill to become an opera star. The same applies to statistical programming – there are some skills that can't be taught but are more at the core of our being.

Problem Solving

Successful statistical programmers are problem solvers. We enjoy a challenge and don't want to quit until we figure it out. We are often driven not only to find a solution, but a "better" solution, such as one that is more efficient or elegant. We impose deadlines and challenges on ourselves that are often tougher than any external ones coming from our manager or study team.

We are often competitive, challenging ourselves to beat others or even just ourselves. However, it is an open and healthy style of competition, because we share our solutions plus learn from and want to apply the solutions of others.

Ability to self-teach

The industry is constantly evolving. New standards are being invented and enhanced, the SAS software package has new features added with each release, and new statistical methods are being developed and applied. We statistical

programmers can't possibly learn up-front everything we'll ever need to use to do our jobs. We instead need to learn as we go, and the fastest way to do this is to be able to teach ourselves.

How we self-teach can differ. Some prefer to collaborate and talk with others. Others are more solitary and research books, papers and websites. Some pour through prior code to pull out sections that can be applied in other places.

The most successful statistical programmers use a combination of these skills, and work toward continual self-improvement. They finish a project, but are never finished with learning.

Altruism

Successful statistical programmers are those who not only are continually learning, but want to share knowledge. This can be demonstrated by presenting or writing papers, but also by simply writing code that is well-organized and documented so that others can easily follow and use it.

SAS programming is done in many industries, not just pharmaceutical. Often without realizing it, many of us have ended up in the pharmaceutical industry because we care about the health and well-being of others. We get excited that, in our own way, we're helping patients. As much work as it is, we enjoy the rush of being in "filing mode" because we know the drug is working and are energized about its potential approval. In fact, we often end up working a lot of unpaid overtime to meet crazy deadlines because we don't want to be the reason a filing is delayed when patients are waiting for the cure.

Most work in the pharmaceutical industry has great ebbs and flows of time commitment. This is unlike some other programming jobs, where you can count on a consistent schedule from week to week. We accept that there might be times, such as during filing preparation, where we need to work a huge number of hours and our social/family life suffers. In fact, we'll take advantage of these busy times to build our business relationships. However, we also make use of the slower periods to reconnect with friends and family outside of work. In other words, successful statistical programmers are comfortable with a schedule that isn't fixed and rigorous, and put work first when maintaining a work-life balance.

ANALYTICAL SKILLS

The ability to communicate, manage time, see the big picture, and lead allow us to analyze a situation and determine what needs to be done. This set of skills, called "analysis skills", includes "soft skills", and are needed in addition to programming skills. In fact, in this industry many of us have job titles that include the term "analyst", not just "programmer".

Communication

A successful statistical programmer must be able to listen and read for understanding. We ask questions to get at the core nuggets of information needed to do our jobs. We can then re-state or rephrase information in a way that others can understand it. We often need to "push" information out to others, such as to keep fellow team members updated on what we're doing. In other cases we need to "pull" information, by asking questions.

Different audience members often need different content even about the same topic. A junior programmer needs to know specifics of what needs to be done, when it needs to be done, and may even need instruction on how to do it. A more senior programmer may just need to know where to find this information and will come to us only with questions or issues. When speaking with our manager, we may only need to give summary information, such as percent completed, and check for understanding.

With today's technology and the fact that teams are becoming more global, communication is more than just face-to-face. Sometimes we need to use phone, email, electronic documents, meetings, or even teleconferences and webcasts to receive or convey information. Here are some examples that might help you decide on which media to use in various situations:

Media	**Good for**
Phone	Checking for understanding
Email	Tracking that a communication took place
Electronic documents	Conveying detailed information
Meetings, teleconferences	Decision making as a group, conveying consistent information to a group
Webcasts	Conveying consistent information to a disperse group

Successful statistical programmers know when to use each type or a combination of media, and check for understanding.

Project management

Managing a project means not only do we need to keep track of what we're doing ourselves, but also what others on the project are doing. One of the biggest challenges of a project manager is influencing without authority, since we are not usually also the manager of the rest of the team and have no formal power.

As a project manager, we need to coordinate activities and prioritize tasks to ensure that all work is done on time. This includes ensuring that staff who are involved in multiple projects are spending the time needed on ours, and helping our staff by pushing back when requests seem unreasonable. We may need to help others keep their focus and thus drive the project forward. Sometimes there is too much or too little work for the team, and we need to recognize when that happens, ask for more resources, or volunteer resources that could be used elsewhere. We have to be experienced enough to notice when timelines are inaccurate or missing necessary tasks, and get those updated.

One of the benefits of project management is that it often involves working with other functional areas. This allows us to learn about what these other groups do and help interaction between groups.

Big-picture focus

Many statistical programmers who are assigned small tasks within a larger project may not understand how their piece contributes to the whole. Their focus is on completing these individual tasks to meet a detailed timeline, and they are primarily concerned about the speed and quality of their programs.

The most successful statistical programmers are those who understand the big picture. We can describe the project in enough detail that our team members can see how their own piece fits into the overall goal. We notice trends, plus look for opportunities to standardize and consolidate tasks, in order to save time and effort. We see what skill set is needed for each task, and assign work so that each person is both sufficiently challenged so as not to be bored plus also able to contribute and feel successful individually.

Leadership

As statistical programmers grow and are given responsibility for larger aspects of a project, often we'll have others working for us on the project in some capacity. Those who delegate well and are able to meet timelines through the work of others are seen as leaders.

Leadership can be described as the ability to get others to work toward a common goal. Sometimes a leader is assigned to a project, and sometimes a leader emerges from the group as the person who motivates the team. Project management, as described above, is a big part of leadership, though even non-project managers can be leaders. As with project managers, the role of a leader doesn't necessarily imply any official authority.

In the past, when we talked about the "leader" of a statistical programming project in the pharmaceutical industry, this implied a largely technical role. A programmer with a lot of technical experience would act as the leader, and lower-level statistical programmers would go to the leader for programming help and advice, and use the leader's code as a template for other programs in the project. Over time, this type of leadership has become a smaller part of the leader's role.

These days, the leader of a statistical programming project in the pharmaceutical industry does more project management and spends less time on complicated programming. Instead of tackling the most challenging programming tasks themselves, a leader often assigns these types of tasks to an expert on the team. No longer does a leader need to be the expert of all technical skills, we just need to know who the experts are so that we can properly assign the work of the project.

This division of labor creates a lot of flexibility on a team. For example, a skilled statistical programmer with a lot of experience in survival analysis can be assigned to perform this type of analysis across a set of projects. This is a more cost-efficient approach than assigning a highly skilled worker to a single project, having them get bored with low-level work, and paying their high rate to also do less-complicated work. Similarly, junior level programmers can be assigned less complicated tasks across several projects, rather than overwhelming them with work they're unprepared to tackle. Many pharmaceutical companies have embraced this breakdown of work, and they make use of high-cost employees, contractors, and consultants only to do complicated work, sending less complicated work to inexpensive locations.

With outsourcing becoming more common, we often find both the high-skilled and low-skilled work outsourced. High-skilled work is typically done by industry experts, often on-site. Frequently the low-skilled tasks are outsourced to inexpensive locations, either in lower-cost areas of the US or in other countries. Leaders are able to balance the time and cost of the project, and we understand when to focus on getting something done quickly versus inexpensively.

Because leader understands the entire project, we will generally assign the tasks and manage the workload, plus help out with whatever programming needs to be done. Whether team members are onsite or outsourced, this delegation process is similar. Leadership positions themselves are not commonly outsourced, because it's much easier for a statistical programmer who is onsite to keep track of changes to the project and liaise with the other functional areas.

Focusing on the group, a leader both get and gives recognition for the entire project. And as a leader, we must recognize that it's more important for the group to succeed than to get credit for any specific task within the project.

SUMMARY

A successful Pharmaceutical Statistical programmer uses a combination of skills: acquired, inherent, and analytical. Acquired skills are those you can learn in order to do the job. Inherent skills are those that come naturally to us and allow us to love what we do. Analytical skills give us the ability to apply what we know to new situations.

Acquired skills include SAS programming, statistics, clinical and industry knowledge, and even an advanced degree. We often learn these skills in traditional classrooms, though many can be picked up online, by reading, or on-the-job. Because the industry is continually evolving, we are never finished with this type of learning. The most successful statistical programmers are constantly learning so that we can keep up with changing technology and standards.

Inherent, unlike acquired, skills are those that we seem to have been born with. These include problem-solving, the ability to self-teach, and altruism. Those of us who have these skills are the most excited about our jobs. While we can get by without these inherent skills, it can be very difficult to be seen as an expert in the industry.

Analytical skills are sometimes called "soft" skills. These include communication, project management, a big-picture focus, and leadership. Many companies offer training in these areas, and courses can be found in both business schools and professional programs. These skills appear to be both acquired and inherent in nature. In other words, we can learn a lot about the topics, but those who are best at it seem to have some natural tendencies in these areas.

While we can get by in our careers without all of these skills, applying the full suite allows us to become truly successful as statistical programmers in the pharmaceutical industry.

REFERENCES

1 SAS Institute offers training all over the world. See http://support.sas.com/learn/index.html for information.

2 Entering "SAS training" in a search engine will bring up a number of schools and private companies who offer training in SAS.

3 The US Food and Drug Administration maintains a website (http://www.fda.gov/) with information for both general consumers and for industry.

4 See the Clinical Data Interchange Standards Consortium's website (http://www.cdisc.org/). They have a "What's New" section on the home page, a tab for "Standards and Innovations", and a tab for "Education and Events".

5 See http://cdiscportal.digitalinfuzion.com/Global%20User%20Networks/default.aspx for information about each of the CDISC global user networks.

6 See the Drug Information Association website (http://www.diahome.org/DIAHome/Home.aspx) for more information. Useful tabs include "Conferences and Meetings", "Training", "Online Learning" and "Publications".

7 See PharmaSUG's website (http://www.pharmasug.org/) for more information.

8 See PhUSE's website (http://www.phuse.com/) for more information. Useful tabs include "Conference" "Single Day Events" and "PhUSE Wiki".

9 See http://support.sas.com/usergroups/index.html for a list of all the users groups, information about upcoming meetings, links to their websites.

10 See American Statistical Association's website (http://www.amstat.org/index.cfm) for more information. Useful tabs include "Meetings & Events", "Career Center", "Education", "Publications", and "Outreach & Initiatives".

11 See Texas A&M's Statistics Department website about their online programs (http://www.stat.tamu.edu/dist/) for more information.

Other papers that may be of interest:

- Smoak, Carey (2009): Empowering SAS® Programmers: The Role of the Manager. http://www.lexjansen.com/pharmasug/2009/ma/ma03.pdf.
- Lafler, Kirk Paul and Shipp, Charles Edwin, (2008): What's Hot, What's Not - Skills for SAS® Professionals. http://www.lexjansen.com/wuss/2008/mcp/mcp05.pdf.
- Koft, Joanna (2006): How to Grow a SAS® Programmer. http://www2.sas.com/proceedings/sugi31/119-31.pdf.

- Shilling, Brian C. and Matthews, Carol (2004): SAS® Programming for the Pharmaceutical Industry. http://www.lexjansen.com/wuss/2004/sas_solutions/i_ss_sas_programming_for_the.pdf.

ACKNOWLEDGEMENTS

The authors would like to thank Mark Matthews and Carey Smoak for reviewing this paper and providing us with useful feedback.

CONTACT INFORMATION

Your comments and questions are valued and appreciated. Authors can be reached at:

Sandra Minjoe
Octagon Research Solutions
585 East Swedesford Rd
Wayne, PA 19087 U.S.A.
sminjoe@octagonresearch.com

Mario Widel
Roche Molecular Systems
4300 Hacienda Dr
Pleasanton CA, 94588 U.S.A.
mario.widel@roche.com