Paper 7940-2016

# Careers in Biostatistics and Clinical SAS® Programming: An Overview for the Uninitiated

Justina M. Flavin, Independent Consultant, San Diego, CA

## ABSTRACT

In the biopharmaceutical industry, biostatistics plays an important and essential role in the research and development of drugs, diagnostics, and medical devices. Familiarity with biostatistics combined with knowledge of SAS® can lead to a challenging and rewarding career that also positively impacts and transforms patients' lives. This paper will provide a broad overview of the different types of jobs and career paths available, discuss the education and skill sets needed for each, and present some ideas for overcoming entry barriers into careers in biostatistics and clinical SAS programming.

## INTRODUCTION

Since 2003, I have been an instructor for courses in SAS Programming and biostatistics/statistics at the University of California, San Diego, Extension Division. My courses are required as part of two professional certificate programs, one in Biostatistics and the other in SAS Programming. These certificate programs and courses attract students with a wide variety of backgrounds, many of whom are interested in pursuing careers in biostatistics and clinical SAS programming. As an instructor and certificate program advisor, I field many questions from students and potential students regarding jobs and career paths, as well as necessary education and skills. These questions form the basis for this paper and will hopefully assist the reader in assessing the feasibility and appropriateness of a career move in this direction.

## INDUSTRY OVERVIEW

### PRODUCT DEVELOPMENT LIFECYCLE

The product development lifecycle of a new drug, biologic or device is a long and expensive process. It typically takes ten to fifteen years - and costs millions if not billions of dollars - to bring a new product to market. The entire lifecycle is complex but can be broken down into three broad stages of Discovery/Research, Clinical Development, and Commercialization.

In the Discovery/Research stage for drugs and biologics, scientists in the lab identify, design, and test thousands of molecules in the search for a promising compound which can potentially treat or prevent a specific disease or condition. Once a candidate compound and a target disease or condition are identified, the compound undergoes further non-clinical testing.Testing includes laboratory experiments and animal studies in which the potential therapeutic effects, safety, and toxicity of the compound are studied and assessed to determine if the compound can proceed into the Clinical Development stage. For medical devices, the Discovery/Research stage starts with the conceptualization of a new device. This is followed by a design and development planning stage. Next, prototypes are built, tested, and further refined prior to defining a final device design. The device then undergoes verification testing when it enters the Clinical Development stage.

In the Clinical Development stage, a research study called a clinical trial is used to expose volunteer human subjects to the compound or device to determine if it is safe and effective. A clinical trial is also commonly used to compare treatments to determine which is better. There are four phases of clinical trials, known as Phases I-IV. In Phase I studies, a small number of usually less than one hundred healthy volunteers are administered the compound. These studies are used to assess a compound's safety, tolerability, and pharmacologic profile. Phase II studies are used to study the safety and efficacy or effectiveness of the compound in subjects having a specific disease or condition. Typically, Phase II studies are larger, enrolling one to two hundred subjects, and the studies are run under very controlled conditions. Phase III studies are referred to as pivotal studies and are also used to assess safety and efficacy in a specific target population. These studies are much larger, enrolling several hundred to thousands of subjects. Upon successful completion of the Phase I-III studies, all of the information from the preclinical studies and clinical trials is compiled into a package of information called a New Drug Application or NDA. The NDA is submitted to various regulatory agencies such as the Food and Drug Administration (FDA) in the US or the European Medicines Agency (EMEA) for review and recommendation. If approval of the NDA is granted by these regulatory agencies, the product moves into the Commercialization stage.

In the Commercialization stage, the manufacturer is allowed to market and sell its product in the countries where regulatory approval has been granted. Once marketed, the manufacturer must continue to monitor and collect information about its product and provide updates to the regulatory agencies regarding the safety of its product. Phase IV studies are initiated in this stage. These are used to continue to collect safety and efficacy information on product usage, to test the use of the product for new indications, and to study cost effectiveness and quality of life issues associated with the product.

### ROLE OF STATISTICS AND PROGRAMMING

In the product development lifecycle, the use of statistics is seen in all three stages. In the Discovery/Research stage, statistical methodology is used to analyze data collected from the laboratory experiments and animal studies. Experimental design methods can be applied to possibly reduce the sample size and the number of experiments or studies needed, thus helping to control costs. Similarly, in the Clinical Development and Commercialization stages, statistical methods are used to properly plan and execute the clinical trials as well as analyze the data collected in these studies. Lastly, in product manufacturing, statistical process control or SPC is used to measure and control variability and ensure product quality.

All of the aforementioned activities generate reams of data, all of which must be collected and stored, manipulated and combined, analyzed and summarized in some meaningful fashion. Programming, in conjunction with statistical software, is used to accomplish these tasks. The FDA does not require or specify the use of any particular programming language or statistical software. However, SAS was adopted by the industry many years ago and a majority of biopharmaceutical and medical device companies currently use SAS, particularly in working with their clinical trials data.

### UNDERSTANDING CLINICAL TRIALS

The conduct and execution of a clinical trial is described in a document called a clinical protocol. The protocol is a written plan which describes the objectives of the study, types of subjects who will be enrolled, schedules of tests and procedures which will be performed, drugs and dosages which will be administered, length of the study, and outcomes that will be measured. The clinical trial is conducted at one or more clinical study sites chosen by the company conducting the clinical trial, also known as the study sponsor. Clinical study sites may include doctor's offices, hospitals, clinics, or outpatient facilities.

Some of the key personnel involved in writing the protocol and conducting the clinical trial include the clinical program manager who serves as a project manager and may have oversight over multiple clinical trials; the medical monitor, a physician who provides medical guidance in the design, execution, and reporting of the clinical trial; the clinical research associate or CRA who is responsible for monitoring the execution of the clinical trial at the clinical study site; the medical writer who compiles and prepares clinical documents, particularly those to be filed with regulatory agencies; the data manager who oversees the design and maintenance of the data base where the clinical trials data are stored; the biostatistician who provides statistical expertise in the clinical trial design and analysis of the data; and the clinical programmer who generates the tables, listings, and figures needed for the analysis and reporting of the study.

Once a clinical trial is completed, a report which is called a Clinical Study Report or CSR presenting the results, findings, and conclusions is prepared and written.

## BIOSTATISTICS RELATED POSITIONS

The vast majority of biostatistics related jobs involve working with clinical trials data, so this discussion will focus primarily on positions in a clinical trials environment. These positions may exist within a sponsor company that is conducting the clinical trials or the positions may be within a Contract Research Organization or CRO, a company to whom the sponsor has outsourced the work.

### BIOSTATISTICIAN/STATISTICIAN – RESPONSIBILITIES

The clinical biostatistician/statistician has the primary responsibility to provide statistical expertise in the design of the clinical trial and in the analysis and associated interpretation of the study data. Other key responsibilities may include writing the statistical methodology section of the study protocol which describes the statistical analyses which are used in the study design and analysis of the clinical trial data collected, determining the correct sample size or number of subjects needed to achieve statistical significance in the context of the study hypothesis being tested, generating the study randomization schedule which is used to randomly assign subjects in the clinical trial to the different treatment groups in the study, writing the Statistical Analysis Plan or SAP which is a document that provides more specific details for performing the statistical analyses described in the study protocol, defining the output

(tables, listings, and figures) needed for the analysis and reporting of the study, reviewing the statistical output and defining any additional ad-hoc output needed, and writing the statistical section of the Clinical Study Report. A more experienced clinical biostatistician/statistician will also serve as the statistical point of contact in meetings, consultations or discussions with the FDA.

The clinical biostatistician/statistician may also review the Case Report Forms (CRFs) which are the instruments used to collect the clinical trials data as well as the data base structures where the data will be stored. A clinical biostatistician/statistician with programming expertise may also write some of the SAS code to perform the statistical analyses, create analysis data sets containing derived variables, or write programming code to validate output produced by others.

Other related jobs in the biopharmaceutical industry for a biostatistician/statistician include opportunities in nonclinical, working with scientists to properly design and analyze preclinical laboratory experiments and animal studies; translational medicine, applying correct statistical methods to ensure that the translational process of a basic laboratory discovery to the clinic for the diagnosis, treatment or prevention of a specific disease is accurate and reliable; pharmacokinetics, providing statistical support for the exploration, analysis, presentation, and interpretation of pharmacokinetic (PK) and phrmacodynamic (PD) data; outcomes research and pharmacoeconomics, applying statistical methodology to determine which treatments work best for which types of patients and to compare the value of one drug or therapy to another.

## BIOSTATISTICIAN/STATISTICIAN – EDUCATION AND SKILLS

To obtain a position as a biostatistician/statistician, an advanced degree (MA/MS/PhD), preferably in biostatistics or statistics, is required. An advanced degree with a strong emphasis in statistics, such as a Masters in Public Health (MPH) with a concentration in biostatistics is also acceptable. Knowledge of statistical software, particularly SAS, is also required. Familiarity with the SAS programming language and the ability to write code in SAS is a very desirable skill that may or may not be required, depending upon the company and position. Written communication skills are very important as a biostatistician/statistician has the responsibility of writing the necessary statistical documents such as the statistical analysis plan, as well as the statistical sections of the clinical protocol and clinical study report. Oral communication skills are also important since the biostatistician/statistician plays a key role on the clinical study team and may also be asked to represent the company in meetings with regulatory agencies such as the FDA.

## CLINICAL PROGRAMMER – RESPONSIBILITES

The clinical programmer has the primary responsibility of writing programs in SAS to generate the output (tables, listings, and figures) needed for the analysis and reporting of the clinical study. Other key responsibilities include reviewing the Case Report Forms (CRFs) and the data base structures where the data will be stored, writing programs known as edit checks to generate output that is used for data cleaning and data review, creating standardized analysis code to be used by all programmers working on the same study, writing programs to perform the statistical analyses and generate ad-hoc output, creating analysis data sets containing derived variables, and writing programs to validate output and data sets produced by others.

In some companies, a distinction is made between clinical programmers and statistical programmers. In these companies, statistical programmers generate all the necessary statistical output needed for the study and work closely with the project statisticians, while clinical programmers work exclusively on non-statistical programming tasks.

Other related roles for clinical programmers include data standards programmer which entails creating data sets in a standardized structure and format for submission to the FDA, and pharmacokinetics programmer which involves working with pharmacologists to produce the data sets needed for performing clinical pharmacokinetics analyses.

## CLINICAL PROGRAMMER – EDUCATION AND SKILLS

To obtain a position as a clinical programmer, a bachelor's degree (BA/BS) is required, preferably in a scientific or quantitative field such as mathematics or statistics. Good quantitative skills and knowledge of elementary statistics is necessary and advanced statistical knowledge is desirable. Extensive knowledge of the SAS programming language is also required. Both written and oral communication skills are also important. A programmer spends much time working solitarily, so it is important to be comfortable with the lack of interaction with other people for much of the work day. Being detail oriented is also a helpful characteristic since programming code must be written to produce output that is consistent and correct.

## PURSUING A BIOSTATISTICS RELATED CAREER

If a biostatistics related position appeals to you, how do you gain the required education and skills to move your career in this direction?

For undergraduate students, consider pursuing a degree in mathematics or statistics. If that is not feasible, some courses in statistics or a minor in one of these areas will be helpful. For graduate students in statistics or biostatistics, courses in survival analysis, multivariate analysis, and experimental design are recommended. For all students, courses providing exposure to the SAS programming language or the SAS statistical procedures are highly recommended.

For professionals considering a career change, one way to explore these areas is through enrollment in professional studies courses and certificate programs. Many universities now offer certificate programs in biostatistics. Quite a few are offered in a fully online format. Most programs consist of regular university courses, many of which are offered through their academic departments. Some of these programs require a formal application process and admission to the program before any courses can be taken.

As mentioned previously, the University of California, San Diego, Extension Division offers two professional certificate programs, one in Biostatistics and the other in SAS Programming. Both of these certificate programs are offered fully online and neither requires formal admission to the program. Students may choose to complete courses prior to enrolling in the certificate program. Courses are offered year round and both certificate programs can be completed in one year or less. As of the Spring 2014 quarter, the Biostatistics certificate cost is $2660, and the SAS Programming certificate cost is $3685. A primary goal in developing these certificates was to create unique programs with courses that would provide exposure to tasks similar to those encountered by people working in jobs in these areas.

The Biostatistics certificate program consists of four courses. Biostatistics covers elementary biostatistics concepts and has a required project in which some simulated real world messy data must be analyzed using statistical techniques learned in the course. A formal written project report is due at the end of the course which must explain the problems studied, analyses used and justification for using these, results presented in graphical and descriptive formats, and conclusions drawn about the results. SAS Programming I covers DATA Step and PROC Fundamentals. By the end of the course, students have written a comprehensive SAS program which reads in data from data files and data sets, combines and transforms the data and produces output which is similar to tables and listings used in the reporting of a clinical trial. Clinical Biostatistics is a survey course presenting statistical methods commonly used in analyzing clinical trials data and the associated SAS statistical procedures used to conduct these analyses. Biostatistical Methods in Clinical Trials is the capstone course which requires students to write the statistical section of a protocol, write a Statistical Analysis Plan, analyze some simulated clinical trial data and write the statistical results section of a clinical study report.

The SAS Programming certificate program consists of six courses. There are two foundation courses, SAS Programming I: DATA Step and PROC Fundamentals and SAS Programming II: Advanced DATA Step Programming, three specialized courses, SAS SQL Programming, SAS Macro Programming, and Output Delivery System and Data Visualization Essentials using SAS. The final course is the SAS Programming Capstone Project. In this course, students will be paired with a mentor. Working with the mentor, the student defines a topic of interest, finds or generates relevant data and uses SAS to demonstrate knowledge of programming techniques and/or data analysis, and writes a technical paper which discusses the associated programming concepts and methodology used in their analyses. The mentors are all SAS experts and experienced presenters at SAS User Group conferences and will assist students in preparing a paper which can be presented to a potential employer as evidence of SAS knowledge and skills or potentially submitted for presentation at a professional conference.

As the Biostatistics certificate program provides exposure to both biostatistical methodology and SAS programming, it is an appropriate choice for those who are unsure of which career path to choose, those who may be considering an advanced degree in biostatistics or statistics, or those who are well versed in SAS programming but not clinical biostatistics. Completion of the SAS Programming certificate is recommended for anyone who wishes to enhance career opportunities by building a strong foundation in the SAS programming language. Those who are completely new to the biopharmaceutical industry and those having limited or no previous exposure to both biostatistics and SAS programming would benefit most by completing both certificate programs.

A similar program, SAS Certification for the Pharmaceutical Industry, is offered at the Philadelphia University School of Continuing and Professional Studies in the Philadelphia, PA area. It is a three month full time intensive program that offers classroom training, an independent project, and internship opportunities on a competitive basis. Over the first eight weeks, students are given an overview of the pharmaceutical industry and specific topics relevant to biostatistics and clinical SAS programming while completing five courses in SAS Fundamentals, SAS Report Writing

and ODS, SAS Programming, SAS Macro Language, and SAS SQL Processing. The final four weeks are spent working on the independent project, applying the topics and concepts learned while working with clinical data to produce the tables and listings for a small clinical study. The program requires an application and admission is competitive. It is offered once a year in the fall and the cost for the 2014 session is $9500.

While the certificate programs offered at UCSD Extension may appear to be more appealing than the traditional program offered at Philadelphia University due to the cost and the convenience of the online format, potential students, especially those who have never taken an online course, should assess their own suitability for online courses. With no scheduled class meetings, having the discipline to set aside the necessary time to complete the coursework and meet the assignment deadlines is of utmost importance in an online course. Traditional courses provide an advantage in promoting personal interaction that cannot be duplicated in online courses. An advantage to online course lectures is that they are recorded, so they can be reviewed and replayed, but a disadvantage is that there is no opportunity to ask questions spontaneously and receive an answer immediately. Additionally, students often report feeling much more isolated and disconnected from their classmates and instructors in online courses. However, online courses have been shown to be as effective as or more effective for learning than traditional courses. Online learning is not for everyone, but for those who are motivated self-reliant problem solvers with good time management skills, online classes can provide a great environment for learning.

## CONCLUSION

For those looking to move their career in the biostatistical direction, there are time and cost considerations that should be taken into account before pursuing any of the educational options presented in this paper. However, future job prospects appear to be good as demand for people with skills in both biostatistics and SAS programming remains high. In 2012, the Bureau of Labor Statistics reported that job growth in these professions was projected to be more than 20% in coming years. In summary, if you are looking for the opportunity to develop and expand your career while collaborating with other talented individuals in positively impacting health for everyone, consider a career in biostatistics and clinical SAS programming.

## REFERENCES


Bureau of Labor Statistics.“Occupational Outlook Handbook”. http://www.bls.gov/ooh/

Flavin, Justina M. (2010). “Strategies for Developing and Delivering an Effective Online SAS® Programming Course”. http://www.wuss.org/proceedings10/Applications/2958_2_APP-Flavin.pdf

Pharmaceutical Research and Manufacturers of America (2014). “Clinical Trials: The Phases of Drug Testing and Approval”. http://www.phrma.org/innovation/clinical-trials

Pharmaceutical Research and Manufacturers of America. “Drug Discovery and Development: Understanding the R&D Process”. http://www.innovation.org/drug_discovery/objects/pdf/RD_Brochure.pdf


## ACKNOWLEDGMENTS


Many thanks to the staff in Instructor Services at UCSD Extension for providing ongoing administrative support, especially Cynthia Hanson and Hilary Lee. I would also like to thank Hugo Villar, Director of Science and Technology, for his guidance and help in developing the specialized certificates in Biostatistics and SAS Programming.


## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at: justina.flavin@gmail.com


Justina Flavin is a consultant providing statistical programming services to clients in the biopharmaceutical and medical device industry. At the University of California, San Diego-Extension Division, Justina is curriculum developer and advisor for the Biostatistics and SAS Programming certificate programs and an instructor for courses in SAS programming and statistics. Justina served as Conference Chair of PharmaSUG ’99 and has taught seminars and been an invited speaker at numerous user group conferences.