

"Breaking Into Statistical Programming: How SAS, R, Python, AI & ML Are Shaping Career Paths"

Kishore Pothuri, Vita Global Sciences, Waltham, MA, USA

ABSTRACT

Breaking into statistical programming without direct industry experience can feel daunting, but there is no single path to success in this evolving field. We spoke with statistical programmers from a variety of backgrounds, mathematics, technology, pharmaceuticals, and agronomy and discovered that transferable skills, technical acumen, and strong problem-solving abilities often outweigh formal credentials.

A degree specifically in statistical programming is not essential; what matters most is analytical thinking, coding proficiency, and an understanding of clinical trial processes. Professionals are drawn to this field for its stability, career growth, and the opportunity to contribute to life-changing healthcare innovations.

Whether transitioning from biostatistics, software engineering, or broader STEM disciplines, building programming expertise in SAS, R, Python, and even emerging technologies like artificial intelligence (AI) and machine learning (ML) can unlock opportunities in clinical programming. This presentation offers practical insights and guidance for aspiring programmers looking to enter industry and thrive in regulatory and research-driven environments.

REAL-WORLD CAREER PATHS INTO STATISTICAL PROGRAMMING

One of the most common misconceptions about statistical programming is that there is only one traditional entry route. In practice, professionals enter this field from a wide variety of academic and professional backgrounds. Conversations with programmers across the industry reveal that adaptability, analytical thinking, and persistence often matter more than the specific degree or initial job title.

Below are a few representative career pathways that illustrate how individuals successfully transition into statistical programming.

TRANSITION FROM INFORMATION TECHNOLOGY TO CLINICAL PROGRAMMING

Many statistical programmers begin their careers in general software development or IT roles. These professionals often already possess strong programming fundamentals, problem-solving skills, and familiarity with data structures.

The key transition typically involves learning clinical trial processes, CDISC standards, and regulatory requirements. Once these domain concepts are understood, technical skills transfer effectively. Several programmers report that mastering SDTM and ADaM concepts, along with gaining exposure to Tables, Listings, and Figures (TLFs), helped them successfully move into clinical research environments.

TRANSITION FROM BIOSTATISTICS OR MATHEMATICS

Individuals with academic backgrounds in biostatistics or mathematics often have strong theoretical knowledge of statistical methods but may initially lack programming experience in regulated environments.

For these professionals, the primary learning curve involves gaining proficiency in SAS or other programming languages and understanding clinical data structures. Many find that their statistical training becomes a strong advantage once they become comfortable with programming and clinical workflows.

TRANSITION FROM FRESH GRADUATES AND EARLY-CAREER PROFESSIONALS

Some statistical programmers begin their careers immediately after completing undergraduate or graduate degrees in life sciences, statistics, or related disciplines.

In these cases, success often depends on building practical skills early, such as learning SAS, practicing TLF generation, and developing sample SDTM or ADaM datasets. Internships, training programs, and mentorship play a particularly important role in helping early-career professionals gain confidence and industry exposure.

TRANSITION FROM OTHER STEM FIELDS

Professionals from fields such as engineering, agronomy, and other scientific disciplines have also successfully transitioned into statistical programming.

These individuals often bring strong analytical thinking and technical problem-solving abilities. With targeted training in clinical trial processes, CDISC standards, and regulatory expectations, they can adapt their existing skills to clinical research applications.

KEY OBSERVATIONS ACROSS CAREER PATHS

Despite different starting points, several common themes emerge among successful statistical programmers:

- Strong fundamentals in programming or analytical thinking
- Willingness to continuously learn new tools and standards
- Practical, hands-on experience gained through real or simulated projects
- Guidance from mentors or experienced colleagues

These patterns reinforce an important message: statistical programming is a field that rewards curiosity, persistence, and continuous learning. While technical tools and industry trends evolve, the core skills that lead to success remain remarkably consistent.

THE EVOLVING ROLE OF STATISTICAL PROGRAMMERS: THE SHIFT TOWARD R, PYTHON, AI AND ML IN STATISTICAL PROGRAMMING

For decades, SAS has been the cornerstone of clinical trial programming, particularly in regulatory submissions. While SAS remains essential, the landscape is changing. Organizations are increasingly incorporating R, Python, and AI-driven approaches into clinical research workflows.

For instance, regulatory agencies, including the FDA, now accept R-based submissions, signaling growing acceptance of open-source tools. Python is also gaining traction for automation, large-scale data processing, and integration with advanced analytics and machine learning models into clinical research.

At the same time, AI and ML are beginning to influence multiple aspects of clinical trials, including:

- Predictive modeling for patient recruitment
- Automated anomaly detection in clinical data
- Risk-based monitoring and quality checks
- Biomarker discovery and exploratory analysis

These changes are not replacing statistical programmers; they are expanding the scope of the role. Programmers who understand both traditional clinical programming and emerging technologies are becoming increasingly valuable.

WHAT HIRING MANAGERS LOOK FOR IN STATISTICAL PROGRAMMERS: STRONG FUNDAMENTALS OVER BUZZWORDS

Hiring managers often prioritize a clear understanding of fundamentals over familiarity with many tools at a superficial level. Candidates who demonstrate a solid grasp of SAS programming concepts, clinical trial workflows, and CDISC standards typically stand out more than those who list numerous technologies without being able to explain how they have applied them.

In recent years, terms such as AI, machine learning, automation, and advanced analytics appear frequently on resumes. While exposure to these areas can be valuable, hiring managers generally look for evidence of practical understanding rather than simply seeing these terms listed. A candidate who can clearly explain how a model was used, how data was prepared, or how results were interpreted is far more compelling than one who mentions these technologies without demonstrating real experience.

Interviewers frequently evaluate how well candidates understand:

- Dataset structures and derivations
- The logic behind Tables, Listings, and Figures (TLFs)
- SDTM and ADaM fundamentals
- Basic statistical concepts used in clinical trials
- The appropriate and practical use of tools, including SAS, R, Python, and emerging technologies

A strong conceptual foundation often predicts long-term success more reliably than exposure to specific tools or trending technologies alone.

BUILDING THE RIGHT SKILL SET

In today's industry, statistical programmers need to build a well-rounded skill set that includes:

- **Proficiency in SAS:** SAS is still a dominant tool in regulatory submissions and clinical data analysis.
- **Expertise in R and Python:** These open-source languages are becoming increasingly popular for statistical modeling, exploratory data analysis, data manipulation, data visualization and ML applications, especially clinical research.
- **Familiarity with AI and ML:** Understanding AI and ML concepts and their applications in clinical trials can provide a competitive edge. Courses on AI and ML applications in healthcare can help bridge the gap between programming and data science.
- **Knowledge of CDISC Standards:** Understanding the Clinical Data Interchange Standards Consortium (CDISC) standards, such as the study data tabulation model (SDTM), analysis data model (ADaM), Define.xml and electronic case report tabulation (eCRT) package submission, is crucial for clinical programmers involved in regulatory compliance and submissions.
- **The Ability to Generate Tables, Listings and Figures (TLFs) for Clinical Studies:** TLF programming remains a core responsibility of statistical programmers and a key skill evaluated during interviews.
- **Experience with Real-World Projects:** Practical experience is often more valuable than theoretical knowledge. Creating sample SDTM datasets, building ADaM datasets, or developing mock TLFs using publicly available data can significantly strengthen a candidate's profile.

PROFESSIONAL CERTIFICATIONS

To maximize your chances of breaking into the field, also consider earning SAS, R & Python certifications. This is a great way to validate your skills and gain credibility in the industry.

Some of the recommended SAS Certifications related to Clinical Programming:

- **SAS Certified Specialist: Base Programming Using SAS 9.4**
Covers data manipulation, DATA step processing, and basic reporting procedures.
- **SAS Certified Professional: Advanced Programming Using SAS 9.4**
Covers advanced DATA step techniques, SQL, macros, and performance optimization.
- **SAS Certified Professional: Clinical Trials Programming Using SAS 9.4**
Focuses on clinical trials data, CDISC standards (SDTM and ADaM), and regulatory reporting.

- **SAS Certified Associate: Programming Fundamentals**
Entry-level certification covering basic SAS programming concepts.
- **SAS Certified Specialist: Statistical Business Analysis Using SAS 9.4**
Covers statistical analysis techniques and procedures used in data analysis.

Some of the recommended R Certifications and Trainings related to Clinical Programming:

- **R Programming Certification – Johns Hopkins University (Coursera)**
Covers data manipulation, statistical analysis, and reproducible research using R. Widely recognized foundational training useful for clinical and biostatistical workflows.
- **HarvardX Data Science Professional Certificate (edX)**
Includes extensive use of R for data wrangling, visualization, and statistical modeling, applicable to clinical research data analysis.
- **PHUSE / CDISC Training Workshops (R for Clinical Reporting – where available)**
Provides practical exposure to using R for clinical trial reporting, tables, listings, and figures (TLFs), and CDISC aligned workflows.
- **Posit (formerly RStudio) Certifications and Tidyverse Training**
Focus on reproducible reporting, R Markdown, and data pipelines increasingly used in regulatory submissions.

Some of the recommended Python Certifications and Trainings related to Clinical Programming:

- **PCAP – Certified Associate in Python Programming (Python Institute)**
Covers Python fundamentals, data structures, and scripting skills applicable to data processing and automation in clinical programming.
- **PCEP – Certified Entry-Level Python Programmer (Python Institute)**
Introductory certification demonstrating basic Python programming skills.
- **IBM Data Science Professional Certificate (Coursera)**
Covers Python, data analysis with pandas, visualization, and basic machine learning—skills increasingly applied in clinical data review and automation.
- **Microsoft Certified: Data Analyst Associate (Python modules included)**
Focuses on data analysis, visualization, and data preparation, including Python-based workflows.

THE ROLE OF MENTORSHIP

One common subject that emerged in our discussions with experienced programmers was the impact of mentorship. Many professionals' credit early guidance from mentors as a key factor in accelerating their careers.

Organizations that invest in mentorship programs often see improvements in employee engagement, retention, and skill development. For aspiring programmers, finding a mentor can shorten the learning curve and provide valuable industry insight.

INTERVIEW PREPARATION

After honing the technical skills above, preparing for interviews is the next step. Here's how to approach it:

- **Practice real-world coding problems:** Sites like PHUSE, PharmaSUG and Kaggle offer great practice materials.
- **Prepare for "soft skill" questions:** Employers look for problem-solving abilities, teamwork and communication skills.

- **Stay updated with industry trends:** Follow AI and ML advancements in clinical research to demonstrate your forward-thinking mindset.

LANDING A STATISTICAL PROGRAMMING ROLE

When looking for a statistical programming role, seek out organizations where you can grow and expand your skill set. Many programmers start at contract research organizations (CROs) or biometrics service providers, where they gain hands-on experience across multiple projects. These roles offer great foundation in clinical trial programming and allow professionals to develop expertise in different therapeutic areas.

CONCLUSION: THE FUTURE OF STATISTICAL PROGRAMMING

The role of statistical programmers is expanding beyond traditional clinical programming. As AI and ML continue to make waves in life sciences, professionals who adapt and learn these advanced technologies will be at the forefront of industry innovation. Companies are increasingly seeking talent with a hybrid skill set, someone who understands regulatory compliance, clinical trial data, statistical methodologies and emerging technologies like AI and ML.

If you're considering moving to a role in statistical programming, now is the time to invest in your skills, expand your knowledge beyond SAS and embrace the power of R, Python, AI and ML. The demand for data-driven professionals in healthcare and clinical research continues to grow, making statistical programming a highly rewarding and future-proof role.

“Statistical programming is no longer just a technical role, it is a career at the intersection of science, data, and innovation.”

RECOMMENDED READING

1.SAS AND CLINICAL PROGRAMMING:

- SAS Institute. *SAS® Certification Prep Guide: Base Programming for SAS 9*. Cary, NC: SAS Institute Inc., latest edition.
- SAS Institute. *SAS® Clinical Trials Programming Using SAS 9.4*. Cary, NC: SAS Institute Inc., latest edition.
- Love, D., & Pittman, R. (2017). *Clinical SAS Programming: A Guide for Beginners*. Springer.

2.CDISC, SDTM, AND ADAM STANDARDS:

- Clinical Data Interchange Standards Consortium (CDISC). *SDTM Implementation Guide*. CDISC, 2025.
- CDISC. *ADaM Implementation Guide*. CDISC, 2025.
- R. A. Califf, et al. (2018). "Standardizing Clinical Data: CDISC Standards for Regulatory Submissions." *Therapeutic Innovation & Regulatory Science*, 52(3), 320–329.

3.R, PYTHON, AND OPEN-SOURCE ANALYTICS IN CLINICAL RESEARCH:

- Gentle, J. E. (2019). *Computational Statistics with R*. Springer.
- McKinney, W. (2023). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. O'Reilly.
- Wickham, H., & Grolemund, G. (2016). *R for Data Science*. O'Reilly Media.

4.AI AND ML IN CLINICAL TRIALS

- Esteva, A., Robicquet, A., Ramsundar, B., et al. (2019). "A Guide to Deep Learning in Healthcare." *Nature Medicine*, 25, 24–29.
- Chen, R., et al. (2020). "Artificial Intelligence in Clinical Trial Design and Analysis." *Drug Discovery Today*, 25(5), 921–927.
- FDA. *Artificial Intelligence/Machine Learning (AI/ML) in Drug Development and Regulatory Review Guidance*. FDA, 2022.

CONTACT INFORMATION:

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

Author Name: Kishore Pothuri

Company: Vita Global Sciences

Address: 281 Winter Street, Suite 100

Work Phone: 781-833-0259

Email: kpothuri@softworldinc.com

Website: www.vitaglobalsciences.com

LinkedIn: <https://www.linkedin.com/in/kishore-pothuri>

Brand and product names are trademarks of their respective companies.