

The Essentials Given – Becoming an Expert of Statistical Programming in Pharma & Health Care Industry

Hongbo Li, Amylyx Pharmaceuticals, INC.

ABSTRACT

In the drug development industry, statistical programming plays an indispensable role for clinical trials or real world data analysis. No matter handling the raw data to submission format data and to analysis-ready data, or creating tables, figures and listings, those deliverables are usually generated by statistical programmers. The responsibilities and duties of statistical programmers also include developing data specifications and programming standards to comply with regulatory requirements and guidelines. Therefore, the good quality of statistical programming work will benefit the entire clinical trial or real world data processing and analysis.

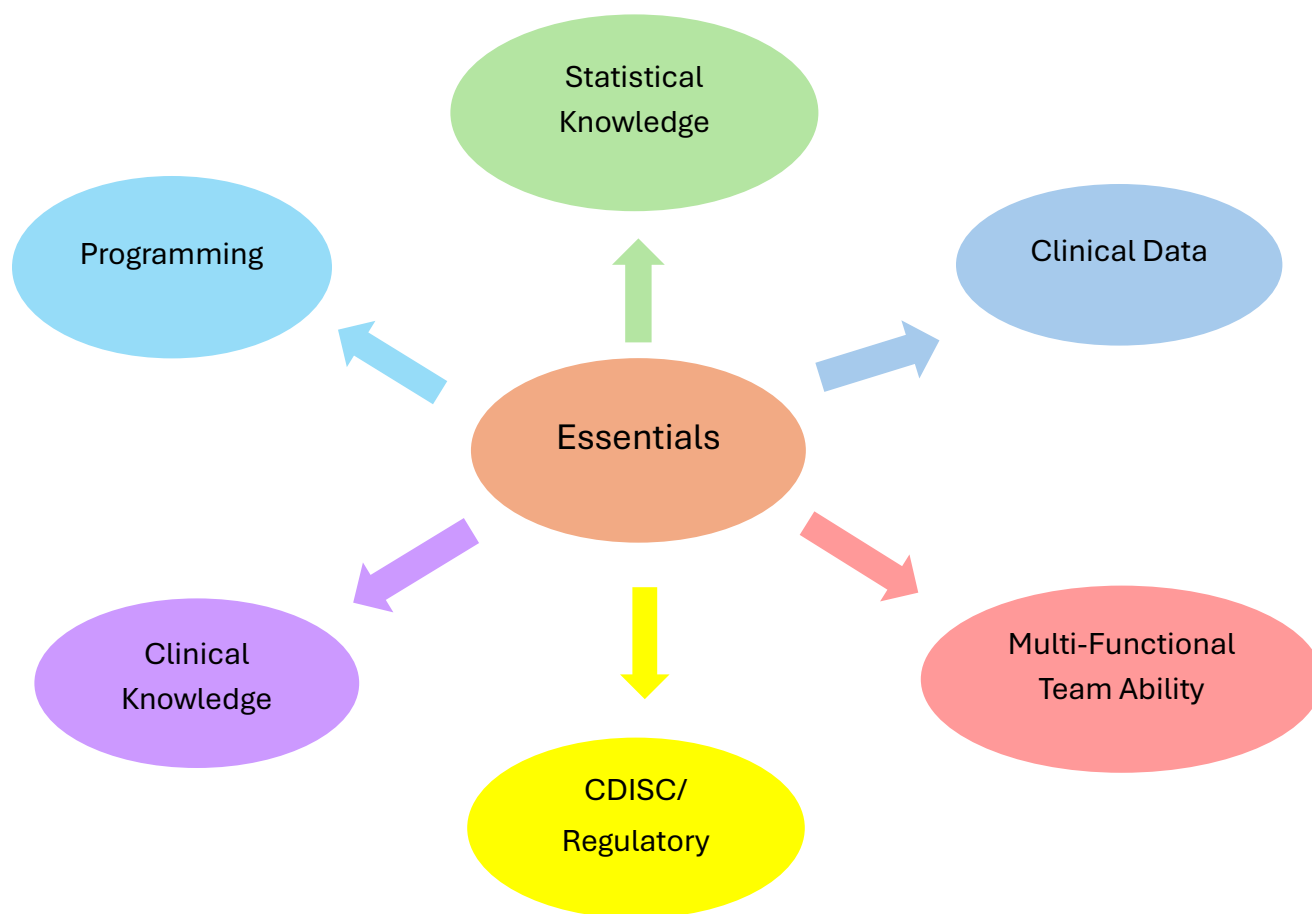
How to become an expert in statistical programming? Based on the author's past more than 15 years industry experience, excellent coding/programming skill (using SAS, R or other software) is not enough at all. Statistical programmers work with biostatistician and data management very closely, and also collaborate with other functional teams like clinical operations, medical writing/clinical development, regulatory affairs, PV (pharmacovigilance), PK/PD (pharmacokinetics/pharmacodynamics) and Biomarker etc. Thus, a good statistical programmer should not only have programming technical, but also statistical knowledge (understanding SAP, mock shells, and statistical modelling etc.), clinical/medical knowledge (for example, knowledge on clinical trials), familiar with CDISC standards and regulatory requirements/compliance checks, having idea about EDC database (electronic data collection system) and performing UAT (user acceptance test) as well as external data collection/transfer process. Also, project management skills and collaborative abilities with cross-functional teams are the essentials to have too for a good statistical programmer. This paper aims to conduct a discussion on those key elements to become an expert in statistical programming area.

Currently, statistical programmers are in high demand due to the growing importance of new drug development in the pharmaceutical and health care industry. Although artificial intelligence technology develops rapidly, the role of statistical programmers is still crucial in the design, development, and execution of data collection, cleaning, manipulation, and statistical analysis. It makes a meaningful contribution to the scientific research for drug development.

INTRODUCTION

In the pharmaceutical and biotech industry, statistical programming plays an important role in drug development. As more younger generations entered this field, I was asked often: How to become an excellent statistical programmer? What knowledge and skills should I achieve beyond programming?

In this paper, I would like to discuss this topic a little deeper in the following six aspects and share some essentials from my 16-years industry experience which may benefit the entry-level and junior statistical programmers, or even senior ones.



PROGRAMMING SKILLS

As a statistical programmer, programming skill is a must for every level of this role. No matter SAS, R, or Python etc. which software you use, you should be able to process data, analyze data and generate corresponding outputs such as datasets, tables, listings and graphs as well as creating macros (simple or complex) as utility tools.

A good statistical programmer should be very familiar with data structure, date format, accessing (read in/write out) data from/to different source/format (excel, csv, xpt etc.), and various statistical procedures/packages. Creating tables, listings and graphs for clinical study reports in RTF or PDF format is very common in statistical programming work, especially for graphs. SAS/GRAPH, ODS graphics and R's ggplot2 package are useful tools to make graphs.

CDISC / REGULATORY KNOWLEDGE

As the pharmaceutical industry goes towards more standardization and harmonization, CDISC standards have been widely adopted within the industry for the past decade. The knowledge of CDASH, SDTM, ADaM, define XML, controlled terminology etc. should be part of statistical programmers' expertise. A

competent statistical programmer should know very well on most recently used version of SDTMIG and ADaMIG (sometimes also the previous version and revision history from different versions) and use them deftly to write SDTM and ADaM specifications as well as creating corresponding SDTM and ADaM datasets.

If a statistical programmer would like to become an expert, he/she should be familiar with requirements and rules for clinical data and analysis outputs from different regulatory agencies, like FDA (US), EMA (EU), PMDA (Japan), MHRA (UK), TGA (AUS), NMPA (China), Health Canada. No matter you have worked with submission packages or not, you may want to have some basic idea about what should be included in the submission packages, how to integrate single studies together for ISS and ISE data packages and do the analysis, what DSUR, PSUR and IB are and what outputs should be generated for those annual reporting periods.

A good statistical programmer should also know about regulatory guidance and compliance, for example, FDA's study data technical conformance guide and technical rejection criteria for NDA, BLA and commercial IND studies. Pinnacle 21 validator (community or enterprise version) is a good tool to use for compliance checks to prepare for submission-ready dataset packages.

COLLABORATION WITH BIOSTATISTICIAN

A biostatistician is the person who makes statistical plan and output/TFL mock shells for a clinical trial. It is the most collaborative functional position with statistical programmer. We, statistical programmers, execute the Biostats' statistical plan, translate it into SAS or R codes, and generate relevant analysis dataset and TFL outputs that Biostats requested.

Therefore, a statistical programmer should have fundamental statistical knowledge to understand the content of SAP and mock shells, statistical models and methods for clinical data analysis, and the output results from different statistical procedures, especially for those complex models (like MMRM, joint model etc.). Statistical programmers should be involved in reviewing SAP and mock shells, providing your expertise to help Biostats on handling missing data, baseline and endpoint definition/criteria, analysis population and critical data analysis.

COLLABORATION WITH DATA MANAGEMENT

Another functional team statistical programmers need to heavily collaborate with is data management. From protocol, CRF design to EDC database build-up, external data collection/transfer, to better understand and present the clinical trial data, statistical programmers should work closely with data managers on data collection and data cleaning activities in the sake of data integrity and accuracy.

An expert in statistical programming must have knowledge of clinical data management, for example, electronic data collection platform, such as Rave, Medidata, Medrio etc., helping with review of data management plan, data validation plan, data transfer specification/plan, performing UAT (user acceptance test), and providing data review listings.

CLINICAL KNOWLEDGE

As statistical programmers are working on clinical trials' data, we need to understand the overall drug development process from research stage to drug approval (phase I to IV), what the difference is between early phase and late phase, how the clinical trials conduct, what clinical monitoring means and so on.

Depending on the investigational products in different therapeutic areas, a statistical programmer should also have some therapeutic area knowledge which may help thorough understanding on the measurement of endpoint assessments/results, and also challenges in a specific therapeutic area such as oncology with corresponding regulatory guidelines.

Medical knowledge is not required for statistical programmers, however, if you have some preliminary concepts on PK/PD (pharmacokinetics/pharmacodynamics) and Biomarker, it will definitely be beneficial for you to work on those related data processing and analysis.

MULTI-FUNCTIONAL TEAM ABILITIES

On some occasions, for example, a small size pharmaceutical or biotech company with no in-house biostats or data manager, a statistical programmer may need to have multi-functional team abilities to work as combination of statistical programming, data management and biostatistics, even with the knowledge of clinical development. Under this situation, you must work outside of biometric area, and directly with clinical operations, medical writing, regulatory affairs, pharmacovigilance, pharmacology etc. those different functional teams, which you need to explain, clarify and mutually understand each other's work. It is something beyond the programming technology and skills, but more about communication, project management skills, conflict resolution, collaborative style and problem-solving strategy.

CONCLUSION

Working in a rapidly changing and evolving world of drug development, statistical programmers should keep a good mix of analytical mindset and comprehensive understanding of data structures and contents. As artificial intelligence and machine learning will come into the pharmaceutical and biotech industry in the future, we, statistical programmers need to be open to the changes and challenges and be willing to jump out of our comfort zone. The above essentials will help us to continue to create value for this industry, by connecting data, performing analysis and providing expertise and knowledge to the team and the whole company.

RECOMMENDED READING

- www.cdisc.org
- www.phusewiki.org
- www.nih.gov/health-information/nih-clinical-research-trials-you/list-registries
- www.ema.europa.eu/en/human-regulatory-overview/marketing-authorisation/marketing-authorisation-guidance-documents
- www.fda.gov/industry/fda-data-standards-advisory-board/study-data-standards-resources

CONTACT INFORMATION

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

Name: Hongbo Li
Enterprise: Amylyx Pharmaceuticals INC.
Phone: 317-219-8433
E-mail: lihongbo1980@gmail.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.