Paper 192-2023

Navigating SAS® and CDISC Certification with Apprenticeship at EDA Clinical

Sarvar Khamidov, EDA Clinical

ABSTRACT

EDA is developing an apprenticeship program to address a shortage of SAS® and Study Data Tabulation Model (SDTM) specialists. The program focuses on achieving SAS and Clinical Data Interchange Standards Consortium (CDISC) certification in a condensed timeframe of 12- to 24-weeks through immersive learning. The primary aim is to equip participants with the essential knowledge and skills for success as a clinical trials programmer.

In this paper, I discuss the effectiveness of this accelerated approach, which is characterized by intensive training sessions, hands-on exercises, group projects, and mentorship, which has spanned more than three months thus far. I share the challenges and achievements during my journey as I evaluate the program's structure. This case study aims to provide guidance for the development of future clinical talent programs to enhance the work force and the industry.

INTRODUCTION

In the pharmaceutical industry, there is a growing demand for people who can transform clinical data. This data is collected on a case report form (CRF) and ultimately submitted for review as XPORT files (XPT). To comply with regulations and explore data efficiently, we conform to standards set forth by the Clinical Data Interchange Standards Consortium (CDISC) including the Study Data Tabulation Model and Define-XML. Additionally, industry specific operating procedures like “double-independent programming” and using standard tools like status trackers and output shells also improve speed and quality of New Drug Applications (NDAs).

Pharmaceutical companies partner with independent contractors or contract research organizations (CROs) to navigate the process of clinical trial data submission. CROs employ a variety of individuals from novices to subject matter experts to facilitate the success of a clinical trial. Their tasks include CRF design, Statistical Analysis Plan (SAP) development, site monitoring, data collection, and data analysis. Programmers help account for all of the data in between.

However, the industry faces challenges in recruiting and training professionals due to the absence of a comprehensive curriculum covering the end-to-end process for clinical trial data. This contributes to resource shortage and delays for biostatistics organizations as the need for statistics programmers is greater than the number of jobseekers (Phares, 2023). Piecemeal training provided on-the-job may overwhelm new candidates and cause attrition and poor quality. Furthermore, an over-emphasis on SAS programming, and not enough on the clinical trials process, may be hampering entry and advancement.

In response, EDA Clinical introduced a “pre-apprenticeship” program to prepare newcomers for actual on-the-job experience or “apprenticeship”. Together, the pre-apprenticeship to apprenticeship path is designed to introduce candidates to the process of reporting and submitting clinical trial data to regulators. In conjunction with coursework, a series of increasingly advanced certifications allows participants to be promoted in a professional setting.

This paper offers my personal account as one of the 12 finalists selected from nearly 500 applicants to participate in the inaugural program.

MY ONBOARDING EXPERIENCE: FROM ON-LINE TO ON-SITE

ON-LINE

I am an international graduate from Penn State University with a bachelor's degree in data science, and I have had internships related to Software Engineering and Data Scientist. As I was graduating May 2023, I was looking for new opportunities in the field of Data Science. I learned of this opportunity from a job listing on-line. It emphasized prior programming experience a chance to help patients at a massive scale. EDA Clinical's website detailed the career paths awaiting successful candidates. Beginning as a student or pre-apprentice, one completes coursework, earns certifications and builds a portfolio that prepares them for a sponsored apprenticeship. Apprentices gain on-the-job experience by working on real, ongoing clinical trials rather than using simulated ones like pre-apprentices do. Plus, as training and experience accumulate, the pay rate is set to rise! The journey seemed promising, so I filled out a brief application.

The next stage of the application process was the virtual interviews. The first interview was an introduction to EDA Clinical, pre-apprenticeship and apprenticeship. This stage included a quiz on the program's content to ensure that candidates are well-informed and know what to expect.

The second interview was more investigative. It was a chance for me to introduce myself to EDA Clinical, sharing my experiences and passion for the field. The Q&A session let us explore what both parties bring to the table.

In the final interview, we discussed the practical aspects of the upcoming program including the start date, pay rate and daily expectations. As someone moving into the city, this was a chance for any final questions I had before taking the plunge.

ON-SITE

Orientation Day

My first day as a pre-apprentice began with an orientation of EDA's facilities. I arrived at work using public transportation, walking past indoor parking as I entered the building. On the ground floor, there is a dining area surrounded by many options for breakfast and lunch. The interior design encourages interaction and group work. It felt like the optimal learning environment, comfortable and convenient, which allowed us to be immersed in the work needed to develop our skills and become clinical trial programmers.

From day one, the goal was to attain the SAS Clinical Trials Programmer Certificate and the CDISC SDTM Certificate within 12 weeks. The SAS Base Programming Certificate is a pre-requisite to earning the Clinical Trials Programming certificate, each expecting 6 and 12 months of relevant experience, respectively. And, to our surprise, the CDISC SDTM Certification exam recommends 3 years of CDISC standards work experience (CDISC, n.d.).

Our first task was to create a syllabus and a schedule to study. This revealed the topics that we would cover and coursework that we would complete on each day. Then, we collected related material including conference papers, samples of study documents, and guidance material (some listed below in Table 1 and in the Recommended Reading Section). We kept a glossary of common terms and acronyms in a notebook that we continue to maintain throughout the program. We organized these items in a shared area by accessing work accounts on our personal laptops.

Table 1. Example of commonly used materials for learning SDTM.

Guidance	Study Data Tabulation Model
	SDTMIG
	SDTMIG-AP
	ADaMIG (1.3)
Conference Papers	An Introduction to SDTM – Part II (Guirk, 2013)
	Best practices for annotated CRFs (Amy, 2020)
Mock Trial Documents	Sample SAPs

	Sample Shells
	Sample CRFs from CDISC

Daily Routine

Each day started with a short meeting around a large conference desk. This time is dedicated to reviewing the progress of the previous day, addressing concerns, and outlining the tasks ahead. After lunch, we reconvened for another 30 minutes to discuss any challenging tasks or coursework. The afternoon sessions were an opportunity to unblock progress, maintain momentum or switch to a new task. Finally, at the very end of each day, we spent our last 15 minutes together organizing our notes, which are instrumental to building our portfolio and tracking our time and progress. Our daily routine played a major role in focusing our teamwork. It also created a dynamic and responsive work environment that can meet our individual needs. This program not only engrained programming skills and industry knowledge but also professional habits.

Coursework

SAS Training and Certification

With EDA's assistance, we enrolled in Coursera's official SAS courses. These courses were the foundation of our knowledge for the SAS Base and Advanced certification exam. It introduced us to SAS syntax and topics like conditional statements, loops and functions but also advanced topics like SQL queries and macros. But we notice that they don't cater to the SAS Clinical Trials Programmer certification. To address this gap, we found conference papers that detail use cases of SAS procedures and data steps. For example, we learned how the retain statement can be used to perform Last-Observation-Carried-Forward or how to use the LAG function to calculate changes from previous visits. We read about how the COMPARE procedure is used to validate results during the double-independent programming process. Throughout our research, we posted relevant questions to our daily agenda to gain additional perspective from our mentors. We also studied any practice exam questions that we could find.

CDISC SDTM Certification Preparation

To prepare for the CDISC SDTM Certification, we developed a comprehensive study guide by comparing exam topics with the details in the SDTM Implementation Guide (SDTMIG). Next, we studied each of the exam topics in detail and presented our findings during the afternoon meetings. We also keep the SDTMIG a few clicks away because we frequently receive “pop questions” in our team’s chatroom.

Still, CDISC recommends test-takers have three years of CDISC experience before attempting the certification exam. This could be particularly challenging for someone like me who is just starting out of college. And, as one of the first pre-apprentices, the program was not yet sponsored by a company with ongoing or completed studies, so clinical trial data was hard to come by.

So, to gain practical experience, we had to simulate study data to standardize, analyze and report. Fortunately, this led to a deep understanding of the clinical trial process. First, we found a sufficiently complex study on clinicaltrials.gov and dissected the study documents. Then, we transformed the CRFs into an electronic format for use in our simulation. Using the schedule of assessments from the SAP, we entered data for each patient visit, completing a profile of each patient. Finally, we manipulated the data to ensure that our analysis would yield the same results found on clinicaltrials.gov. Now, we have raw data to transform into tables, figures and listings (TFLs) using SDTM and Analysis Data Model (ADaM) datasets. In the future, we will create “dirty data” by including partial or missing dates, unscheduled visits, and more. Our defensive programming techniques should limit the changes necessary when we re-run the reports and the QC process with this new data.

Projects
Standard Operating Procedures

We developed standard operating procedures and related tools to further simulate on-the-job experiences. For example, we established a mock deliverable tracker to list all the programs that we were developing and quality-checking (QC). We designed a workbook to collect specifications for every dataset in the tracker. We maintained TLF shells to specify the titles, footnotes and formatting options that were expected in each output. These tools underscored the importance of quality, traceability and timeliness in the day-to-day work of a clinical trials programmer.

Annotations, Specifications and Output Destinations

We annotated the fields on the CRFs (Figure 1) and annotated the columns on the TLF shells. These CRF annotations included the names of SDTM domains and variables, aiding in the specification of standard datasets that conform to SDTM. The TLF annotations help us map the SDTM variables to the columns in the shells. TLF annotations also helped us identify the need for ADaM variables to produce the final result or display. We develop specifications for ADaM datasets before double-programming them, too. Both the SDTM and ADaM specifications improve as we resolve discrepancies and adapt our logic. Finally, we load the QC-complete ADaM datasets into the REPORT procedure to produce reports. Using the Output Delivery System (ODS), we can produce our reports as HTML, PDF, RTF, Excel and more filetypes. Later, we plan to use Statistical Graphics (SG) procedures, like SGPLOT and SGPANEL, to produce figures and dashboards.

Figure 1. Sample Mock Annotated CRF

Double-Independent Programming

My role also introduced me to the meticulous double-independent programming process. This involved at least two individuals separately programming the same deliverable in SAS or an alternative programming language, with a focus on ensuring that the results were equivalent based on the same dataset specifications and TLF shells. With access to Python, R and SAS OnDemand, we began to understand the challenges and nuances associated with preserving metadata and traceability.

Because we could not read and write to a shared location using SAS OnDemand for students and learners, we transferred files to one another to perform our comparison using PROC COMPARE. Nevertheless, the discussions that ensued played a pivotal role in identifying logical inconsistencies and refining the specifications that we shared. We resolved datasets and variables with the most dependencies, first, hoping that downstream differences would also be fixed. These rigorous processes safeguarded the accuracy of the results and the integrity of the simulated submission.

RESULTS AND REFLECTIONS

Starting this journey, I felt a deep sense of accomplishment and gratitude for the knowledge and experiences I've quickly amassed. Although we didn't achieve our audacious 12-week goal, the progress we've made is notable. Personally, I earned two of the three available SAS certifications: SAS Base and SAS Clinical Trials Programmer. The SDTM and SAS Advanced certifications are still on my to-do list.

As shown in Figure 2, we collected and categorized our time by designing data entry form, an independently double programming a table to recount our weekly tasks.

		SDTM		ADaM		Table			Listing			Figure				
Deliverables	Role	Created	Updated	Created	Updated	Unique	Repeat	Update	Unique	Repeat	Update	Unique	Repeat	Update	Others	Comments
EDA Clinical	Main	0	0	0	0	2	0	0	1	0	0	0	0	0	37	2023-09-18: CRF generation and annotation 2023-09-19: Created Tables and Listings 2023-09-19: Annotated more CRFs 2023-09-19: Learned about FDA submission process 2023-09-20: APERIOD Macro 2023-09-20: Medical devices, associated persons review 2023-09-20: more aCRFs that were necessary 2023-09-21: EVEN MORE CRFs 2023-09-21: Patient Profile 2023-09-21: Patient Profile Revision 2023-09-22: Researched dataset-json 2023-09-22: Created powerpoint presentation on json, REST API, dataset-json 2023-09-22: Table on when to use domains 2023-09-22: Last CRF annotation, on snoring phenotype questionnaire 2023-09-22: Finished macro, revised based on input
	QC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
SAS Certifications	Main	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	QC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
SDTM Certification	Main	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	QC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Total															40	

Figure 2: Weekly Task Report

Reflecting on the program, while there's a wealth of knowledge provided, there's always room for improvement. While the advanced topics, including SQL, SAS Macros and Dataset-JSON, were beneficial, strictly focusing on exam topics required for certification might have been more strategic. The curriculum could also benefit from more real-world application exercises that mirror the challenges faced in actual clinical trials. This could bridge the gap between theoretical understanding and practical application, ensuring participants are even better prepared for the demands of the clinical research sector.

Another area for improvement lies in identifying candidates who have a natural flair for programming, a passion for clinical trials and shared long-term goals. To do so, EDA should consider new sourcing strategies including partnerships with universities, coding bootcamps, or even hosting hackathons to attract top talent. EDA should also prepare sponsors to interview pre-apprentices as they earn the proper certifications and build their portfolios.

Lastly, while the program features several strategies to keep participants focused and engaged, the demands of certification and the speed of the curriculum can lead to burnout. To counter this, EDA must cater to different learning styles. Also, regular sessions with a variety of mentors could offer new perspectives, keep motivation high and ensure a continuous zest for learning.

NEXT 12 WEEKS

We established short-term goals to attain our long-term goal of supporting an on-going or upcoming clinical trial:

- Deliverables: We will output more mock deliverables, focusing on developing and validating SDTM and ADaM datasets, TLFs, and, in due course, define.xml and XPT files.
- CDISC Implementation: We adhere to the CDISC standards prescribed by the FDA standards catalog (FDA, 2023) to create submission-ready study data and reports for our mock trial. I currently use SAS, Python, and R to perform independent programming, which ensures the quality of our datasets and specifications.
- SDTM Certification: I am leveraging the hands-on experience from our mock trial to prepare for the certification exam. Fortunately, the mock trial requires the use of Associated Persons and Medical Device domains which helps us build our knowledge of SDTM
- Conference Presentation: I am presenting at an upcoming conference to share my experience, assist others, and amplify my newfound skills.
- Program Sponsor: A sponsor can play a vital role in providing additional resources, mentorship, and opportunities for the next cohort of pre-apprentices seeking on-the-job experience.

CONCLUSION

As I look forward to the future, I am excited about the impact we can make in clinical trials as an industry. The knowledge and skills acquired during this training program at EDA Clinical, previous internships and

at Penn State have positioned me to contribute significantly to the advancement of clinical research processes. Furthermore, I am hoping for opportunities that entail more advanced analytical skills, including elements of machine learning. Embracing this transformative journey has not only enriched my professional capabilities but has also instilled a commitment to supporting the next cohort as they navigate the complexities of the field. I am eager to continue my journey, applying and expanding my expertise in the dynamic landscape of clinical research and data management.

REFERENCES

CDISC. (2021). Version 2.0. Study Data Tabulation Model. https://www.cdisc.org/standards/foundational/sdtm.

CDISC. (2022). Version 3.4. Study Data Tabulation Model Implementation Guide: Human Clinical Trials. https://www.cdisc.org/standards/foundational/sdtmig.

CDISC. (2013). Version 1.0. Study Data Tabulation Model Implementation Guide: Associated Persons. https://www.cdisc.org/standards/foundational/sdtmig/sdtmig-ap-v1-0.

CDISC. (2018). Version 1.1. Study Data Tabulation Model Implementation Guide for Medical Devices. https://www.cdisc.org/standards/foundational/medical-devices-sdtmig/sdtmig-medical-devices-v1-1.

CDISC. (2021). Version 1.3. Analysis Data Model Implementation Guide. https://www.cdisc.org/standards/foundational/adam.

U.S. Food & Drug Administration. (2023). Data Standards Catalog. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/data-standards-catalog

Philips Respironics. (2019, February 15). In Home Assessment of Three Anti-Snoring Devices, a Cross Over Study. https://clinicaltrials.gov/study/NCT03333876?a=3

Guirk, J. M., IS03: An Introduction to SDTM – Part II., https://www.lexjansen.com/phuse/2013/is/IS03_ppt.pdf. PowerPoint Presentation.

Amy, G. (2020). Best practices for annotated CRFs., PharmaSUG. https://www.lexjansen.com/pharmasug/2020/DS/PharmaSUG-2020-DS-248.pdf.

Matthews, C. I. & Shilling, B. C., (2008). Validating Clinical Trial Data Reporting with SAS. SAS Institute

Phares, C. (2023a, August 2). Exploring the skills gap in statistical programming organizations. PharmTech. https://www.pharmtech.com/view/exploring-the-skills-gap-in-statistical-programming-organizations

ACKNOWLEDGMENTS

I would like to extend a special thanks to James Joseph, Justin Man, and Rachel Sieber at EDA Clinical who have shaped both my experiences and this paper.

RECOMMENDED READING

Holland, C., & Shoshak, J. Implementing CDISC Using SAS®

Matthews, C. I., & Shilling, B. C. (2008). Validating Clinical Trial Data Reporting with SAS. SAS Institute

Mandy, B., Otis, E., Stephen, T. (2015). CRO, TLF, SOP? OMG!: A Beginner's Guide to the Clinical Research Organization. PharmaSUG. https://www.pharmasug.org/proceedings/2015/CP/PharmaSUG-2015-CP04.pdf

Guirk, J. M., IS03: An Introduction to SDTM – Part II., https://www.lexjansen.com/phuse/2013/is/IS03_ppt.pdf. PowerPoint Presentation.

Kristin, K. & Michael, B.. (2023). 'TUFF' LOVE: SDTM Pain Points Explained!. PharmaSUG, US. https://www.pharmasug.org/proceedings/2023/DS/PharmaSUG-2023-DS-186.pdf

https://www.lexjansen.com/wuss/2009/cdi/CDI-Wood.pdf

Getting Started with SAS® Programming. (n.d.). Coursera. https://www.coursera.org/learn/sas-programming-basics

SAS Advanced Programmer. (n.d.). Coursera. https://www.coursera.org/professional-certificates/sas-advanced-programmer

Chakravarthy, V. (2018). The Anatomy of Clinical Trials Data: A Beginner's Guide. SAS. 2023, https://support.sas.com/resources/papers/proceedings18/1759-2018.pdf

CONTACT INFORMATION

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

Sarvar Khamidov

EDA Clinical LLC

Sarvar.Khamidov@EDAClinical.com

https://www.linkedin.com/in/sarvarkhamidov/