PharmaSUG 2025 - Paper LS-156

Essential Elements for an Effective New-Hire Training Handbook

Ingrid Shu, Xinhui Zhang, Merck & Co., Inc., Rahway, NJ, USA

ABSTRACT

New hires entering the pharmaceutical industry possess the technical skills required to excel, but if industry-specific concepts and workflows are not adequately introduced during onboarding, the learning curve may remain steep. This results in new hires struggling to unlock their full potential and experiencing hardship in deriving meaning from their work. To reduce growing pains for both mentors and new hires, teams can adopt an informal training handbook that pays special attention to their team's unique processes. This handbook does not replace official resources; instead, it complements them, using simplified language and presentation specifically tailored for programmers transitioning into their role for the first time.

To facilitate effective onboarding, the handbook should be aligned with 3 key principles: centralization, relevance, and foundational knowledge. This paper delves into the importance of these best practices and showcases examples from a SharePoint site designed for new hires in the Early-Stage Oncology Analysis & Reporting statistical programming team we belong to.

Developing accessible new-hire training materials is a rewarding endeavor for the entire team. When hiring junior employees, managers can be confident that this tool will effectively prepare them for their roles. For junior staff, this resource reduces confusion and provides clear direction, fostering empowerment and lowering turnover rates. Ultimately, both senior and junior programmers experience improved performance and productivity, as less time will be spent on training. Not only does it smoothly orient new hires, but it also serves as an ongoing valuable reference even as they grow more accustomed to their role.

INTRODUCTION

New employees, particularly those venturing into the field of statistical programming for the first time, are stepping into a realm resembling a foreign country with its own distinct language. These languages include not just SAS and R, but also industry-specific terminology, encompassing concepts like ADaM, TLFs, specifications, and deliverables. Those with prior experience in similar roles at different organizations may find it easier to acclimate, as they have encountered related “dialects” in their previous workplaces. In contrast, entirely new hires must learn to navigate this foreign landscape from the ground up, often lacking any familiarity with its intricate vocabulary.

It is thus crucial to think from the perspective of a brand new hire when developing training materials. While concepts like "database lock" may seem self-explanatory to the experienced programmer, it is important to first provide definitions and purpose before jumping ahead to details and instructions. A “no assumptions” mindset should guide this process.

Many departments offer an extensive collection of resources which contain content generalizable to the entire department. When first starting out, new hires may find it challenging to navigate these materials and identify which resources and sections within lengthy documents are pertinent to their team and their roles. Furthermore, different teams within the department may operate with varying timelines, formats, and procedures. As such, it is essential for each team to share clear documentation of their group's own unique workflows with their members. Thus it is helpful to provide a deliverable roadmap that is specific to your team and role. This roadmap outlines each deliverable step/task along with a (1) simple definition, (2) simple instructions on how your team does it, and (3) hyperlink to official company resources that have in-depth, additional information.

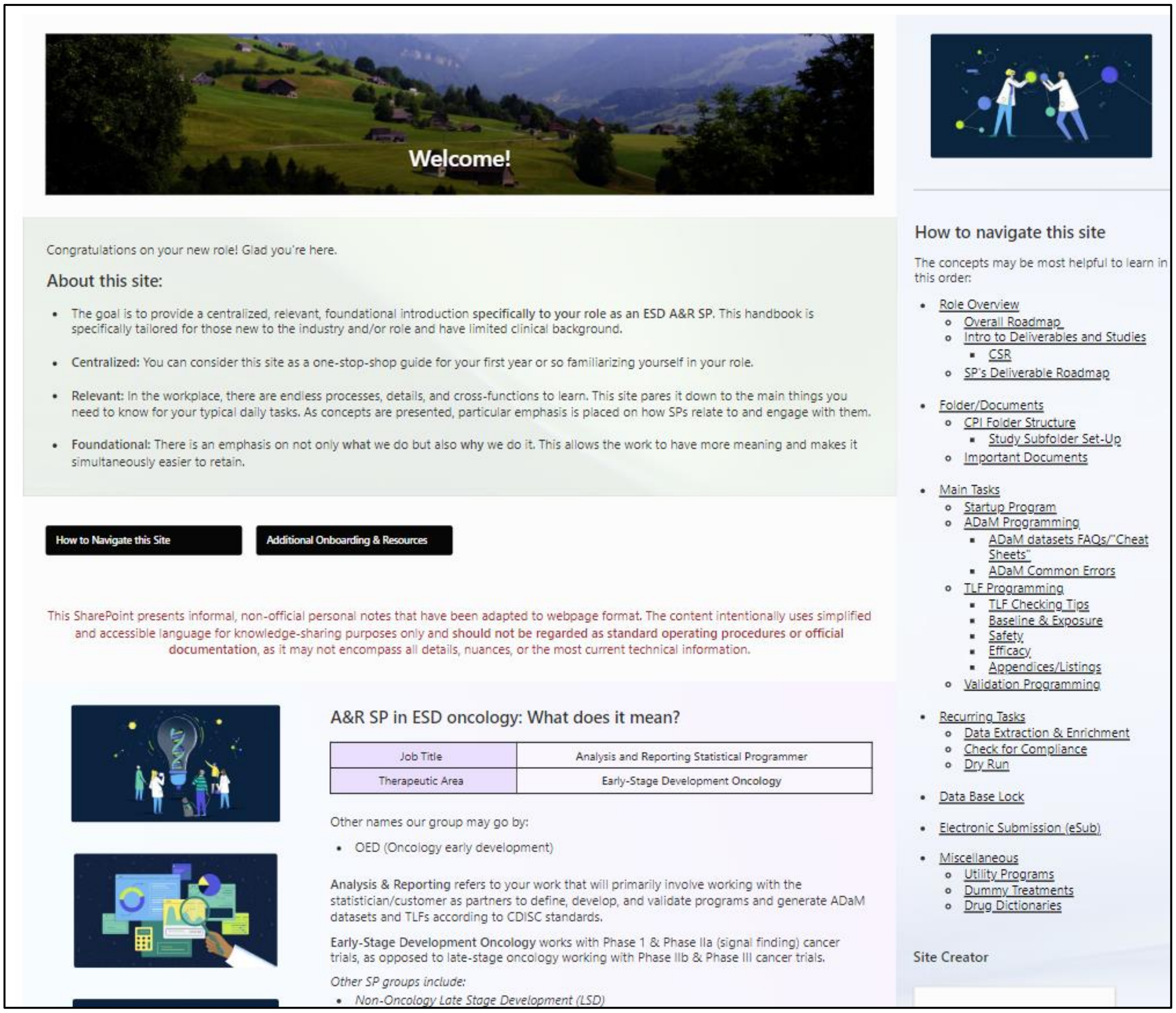
The primary questions that we want new hires to be able to answer are: (1) What should I do? (2) Why do we do it? (3) How to do it? By adhering to the guiding principles of centralization, relevance, and foundational knowledge, you can craft training materials that effectively address these questions.

CENTRALIZATION

It is essential to have a centralized hub containing all relevant information for new hires, ensuring convenient, time-saving access to reference materials. Without a centralized spot for your team's new hire resources, it is possible that important material could be overlooked without the new hires even realizing it. This centralized hub can provide links to supplementary resources, but new hires can be rest assured that they will be able to find what they need in this one spot.

One way to think about centralized new hire training material is that it can serve as a resource for Frequently Asked Questions, where new hires can reference clearly documented information on processes they will inevitably require assistance with as they are learning. Juxtapose this with a system lacking clear documentation: each individual new hire will have to approach their mentor each time they have a procedural question like "how do I perform a data extraction?" or a conceptual question like "why is it important to have both a development and production environment?" The mentor must then invest time to explain each procedure and depend on the new hire to either retain the information or record it in their personal notes. This reliance can result in disorganization and uneven learning experiences among employees.

One effective platform to centralize all your information is SharePoint, which offers flexibility and organization for presenting information through pages and subpages.



Welcome!

Congratulations on your new role! Glad you're here.

About this site:

- The goal is to provide a centralized, relevant, foundational introduction **specifically to your role as an ESD A&R SP**. This handbook is specifically tailored for those new to the industry and/or role and have limited clinical background.
- **Centralized:** You can consider this site as a one-stop-shop guide for your first year or so familiarizing yourself in your role.
- **Relevant:** In the workplace, there are endless processes, details, and cross-functions to learn. This site pares it down to the main things you need to know for your typical daily tasks. As concepts are presented, particular emphasis is placed on how SPs relate to and engage with them.
- **Foundational:** There is an emphasis on not only **what** we do but also **why** we do it. This allows the work to have more meaning and makes it simultaneously easier to retain.

This SharePoint presents informal, non-official personal notes that have been adapted to webpage format. The content intentionally uses simplified and accessible language for knowledge-sharing purposes only and **should not be regarded as standard operating procedures or official documentation**, as it may not encompass all details, nuances, or the most current technical information.

A&R SP in ESD oncology: What does it mean?

Job Title	Analysis and Reporting Statistical Programmer
Therapeutic Area	Early-Stage Development Oncology

Other names our group may go by:

- OED (Oncology early development)

Analysis & Reporting refers to your work that will primarily involve working with the statistician/customer as partners to define, develop, and validate programs and generate ADaM datasets and TLFs according to CDISC standards.

Early-Stage Development Oncology works with Phase 1 & Phase IIa (signal finding) cancer trials, as opposed to late-stage oncology working with Phase IIb & Phase III cancer trials.

Other SP groups include:

- *Non-Oncology Late Stage Development (LSD)*

How to navigate this site

The concepts may be most helpful to learn in this order:

- Role Overview
  - Overall Roadmap
  - Intro to Deliverables and Studies
    - CSR
  - SP's Deliverable Roadmap
- Folder/Documents
  - CPI Folder Structure
    - Study Subfolder Set-Up
  - Important Documents
- Main Tasks
  - Startup Program
  - ADaM Programming
    - ADaM datasets FAQs/"Cheat Sheets"
    - ADaM Common Errors
  - TLF Programming
    - TLF Checking Tips
    - Baseline & Exposure
    - Safety
    - Efficacy
    - Appendices/Listings
  - Validation Programming
- Recurring Tasks
  - Data Extraction & Enrichment
  - Check for Compliance
  - Dry Run
- Data Base Lock
- Electronic Submission (eSub)
- Miscellaneous
  - Utility Programs
  - Dummy Treatments
  - Drug Dictionaries

Site Creator

Figure 1 Landing page of New Hire Training Site

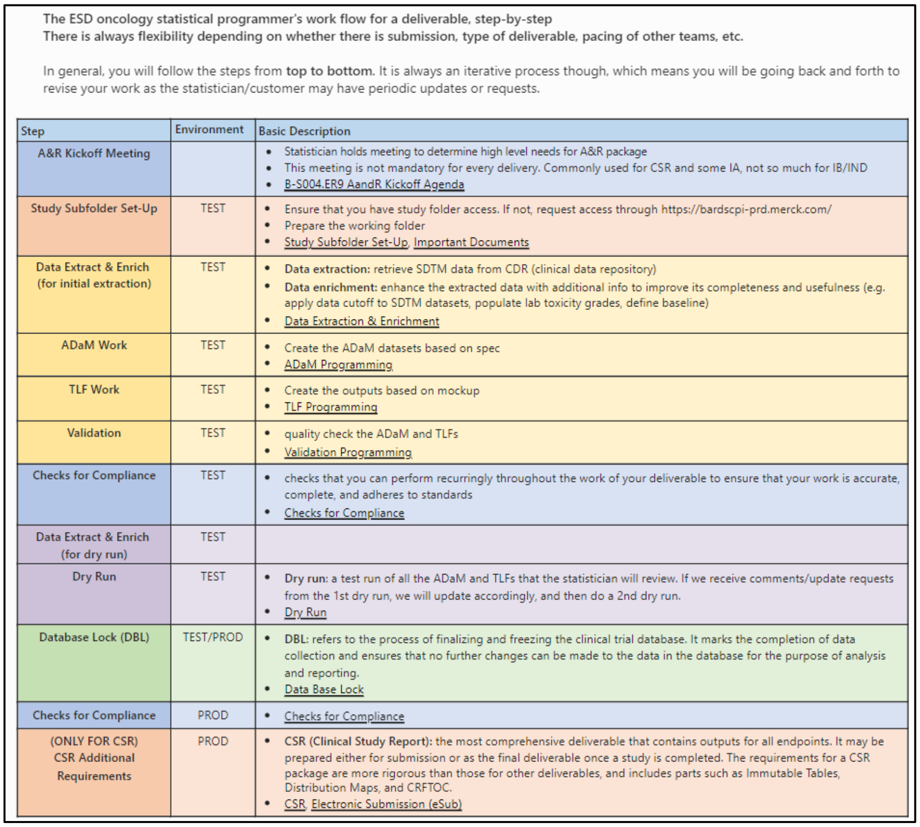
RELEVANCE

New programmers may feel inundated when they receive information from various sources, such as monthly staff meetings, lengthy protocols, mandatory compliance trainings, and guest seminars. The new hire is exposed to much of this information before they even have a chance to grasp the basics of concepts that are at the crux of their daily work. While all this information is helpful to be aware of in due time, it is not essential for them to know until they have first established their foundational knowledge. Without clear guidance to direct their attention to critical concepts first, new programmers may unintentionally expend their energy and efforts on areas that are not directly relevant to their daily work.

Numerous companies maintain a page that links SOPs by department. SOPs are crucial since they ensure compliance and standardization of work across the organization. However, a library of SOPs is not a substitute for training material. A more effective approach for onboarding a new programmer is to develop a roadmap that sequentially outlines the most common steps in their role's deliverables. Within each step, this roadmap should (1) provide a simple description of the step and (2) hyperlink the applicable SOPs that provide relevant information and instructions.

This team-specific deliverable roadmap is beneficial for the new employee as it helps them gain an early understanding of their responsibilities, orient themselves in which step of the process they are in, and determine the areas where they should concentrate their efforts in learning. Furthermore, the combination of a high-level description paired with low-level SOPs ensures that they grasp fundamental concepts while also adhering to the standards.

In some cases, specific deliverable steps may not have dedicated SOPs due to the team's unique approach, or the instructions may be scattered across multiple lengthy documents. It is helpful to extract and consolidate the most relevant information from these disparate sources into a single, concise outline of the instructions for completing the task. This type of documentation offers the new hire a straightforward, accessible reference instead of sending them on a scavenger hunt for something they may not even recognize.



The ESD oncology statistical programmer's work flow for a deliverable, step-by-step
There is always flexibility depending on whether there is submission, type of deliverable, pacing of other teams, etc.

In general, you will follow the steps from **top to bottom**. It is always an iterative process though, which means you will be going back and forth to revise your work as the statistician/customer may have periodic updates or requests.

Step	Environment	Basic Description
A&R Kickoff Meeting		• Statistician holds meeting to determine high level needs for A&R package • This meeting is not mandatory for every delivery. Commonly used for CSR and some IA, not so much for IB/IND • B-S004.ER9 AandR Kickoff Agenda
Study Subfolder Set-Up	TEST	• Ensure that you have study folder access. If not, request access through https://bardscpi-prd.merck.com/ • Prepare the working folder • Study Subfolder Set-Up, Important Documents
Data Extract & Enrich (for initial extraction)	TEST	• **Data extraction:** retrieve SDTM data from CDR (clinical data repository) • **Data enrichment:** enhance the extracted data with additional info to improve its completeness and usefulness (e.g. apply data cutoff to SDTM datasets, populate lab toxicity grades, define baseline) • Data Extraction & Enrichment
ADaM Work	TEST	• Create the ADaM datasets based on spec • ADaM Programming
TLF Work	TEST	• Create the outputs based on mockup • TLF Programming
Validation	TEST	• quality check the ADaM and TLFs • Validation Programming
Checks for Compliance	TEST	• checks that you can perform recurringly throughout the work of your deliverable to ensure that your work is accurate, complete, and adheres to standards • Checks for Compliance
Data Extract & Enrich (for dry run)	TEST	
Dry Run	TEST	• **Dry run:** a test run of all the ADaM and TLFs that the statistician will review. If we receive comments/update requests from the 1st dry run, we will update accordingly, and then do a 2nd dry run. • Dry Run
Database Lock (DBL)	TEST/PROD	• **DBL:** refers to the process of finalizing and freezing the clinical trial database. It marks the completion of data collection and ensures that no further changes can be made to the data in the database for the purpose of analysis and reporting. • Data Base Lock
Checks for Compliance	PROD	• Checks for Compliance
(ONLY FOR CSR) CSR Additional Requirements	PROD	• **CSR (Clinical Study Report):** the most comprehensive deliverable that contains outputs for all endpoints. It may be prepared either for submission or as the final deliverable once a study is completed. The requirements for a CSR package are more rigorous than those for other deliverables, and includes parts such as Immutable Tables, Distribution Maps, and CRFTOC. • CSR, Electronic Submission (eSub)

Figure 2 Roadmap tailored to Merck's Early-Stage Oncology A&R SP deliverable work

Further examples of "relevance" as a guiding principle include instructing new hires on how to read SOPs, emphasizing the sections pertinent to their specific roles, and highlighting the most critical aspects of the protocol for statistical programmers. “Relevance” is essential for productive learning.

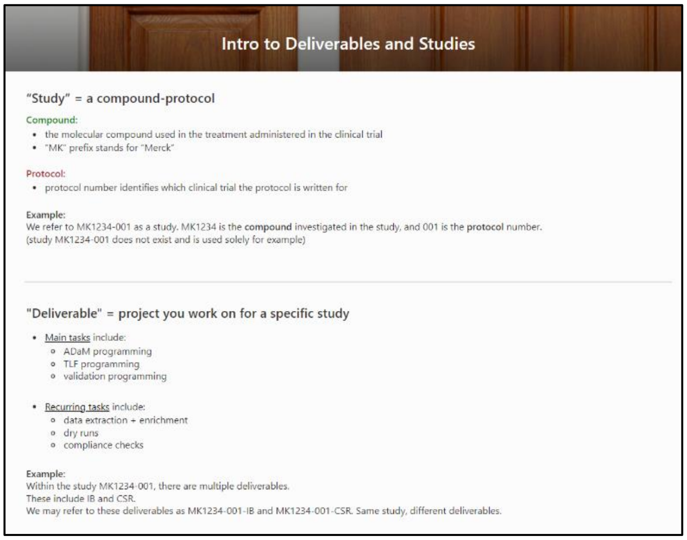
FOUNDATIONAL KNOWLEDGE

Experienced statistical programmers are as familiar with deliverable tasks as a master pastry chef is with their own apple pie recipe. Imagine you are teaching someone to bake an apple pie, but your student has never seen or tasted one before. You, the instructor, may jump directly into the detail of each recipe step, intricately explaining crust crimps and spice ratios. After all, you have been making and eating pies for years, so you cannot imagine not knowing the basics inside and out. By the end of the session, the student may follow the recipe steps perfectly, but without understanding how each ingredient affects the overall flavor and texture of the pie. Likewise, new employees need to understand the rationale behind their tasks, not just the steps to follow. Knowing the “why” and greater goal behind tasks helps employees not only remember procedural steps, but also helps them feel more engaged and connected with their work.

Here are three examples of how foundational knowledge can be implemented in new hire training material.

EXAMPLE #1: PLAIN AND SIMPLE

Figure 3 is from an introductory page to Deliverables/Studies. Terminology and formatting are introduced simply and clearly. Included is a table that explains some of the most common deliverable types. This is important to distinguish early on, so new hires know what different deliverables entail, rather than perceiving them as random acronyms.



Intro to Deliverables and Studies

“Study” = a compound-protocol

Compound:
- the molecular compound used in the treatment administered in the clinical trial
- "MK" prefix stands for "Merck"

Protocol:
- protocol number identifies which clinical trial the protocol is written for

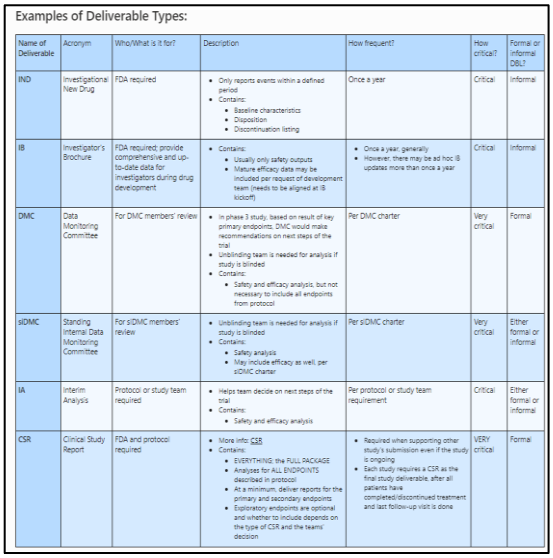
Example:
We refer to MK1234-001 as a study. MK1234 is the **compound** investigated in the study, and 001 is the **protocol** number.
(study MK1234-001 does not exist and is used solely for example)

"Deliverable" = project you work on for a specific study

- Main tasks include:
  - ADaM programming
  - TLF programming
  - validation programming

- Recurring tasks include:
  - data extraction + enrichment
  - dry runs
  - compliance checks

Example:
Within the study MK1234-001, there are multiple deliverables.
These include IB and CSR.
We may refer to these deliverables as MK1234-001-IB and MK1234-001-CSR. Same study, different deliverables.



Examples of Deliverable Types:

Name of Deliverable	Acronym	Who/What is it for?	Description	How frequent?	How critical?	Formal or informal DBL?
IND	Investigational New Drug	FDA required	• Only reports events within a defined period • Contains: • Baseline characteristics • Disposition • Discontinuation listing	Once a year	Critical	Informal
IB	Investigator's Brochure	FDA required; provide comprehensive and up-to-date data for investigators during drug development	• Contains: • Usually only safety outputs • Mature efficacy data may be included per request of development team (needs to be aligned at IB kickoff)	• Once a year, generally • However, there may be ad hoc IB updates more than once a year	Critical	Informal
DMC	Data Monitoring Committee	For DMC members' review	• In phase 3 study, based on result of key primary endpoints, DMC would make recommendations on next steps of the trial • Unblinding team is needed for analysis if study is blinded • Contains: • Safety and efficacy analysis, but not necessary to include all endpoints from protocol	Per DMC charter	Very critical	Formal
siDMC	Standing Internal Data Monitoring Committee	For siDMC members' review	• Unblinding team is needed for analysis if study is blinded • Contains: • Safety analysis • May include efficacy as well, per siDMC charter	Per siDMC charter	Very critical	Either formal or informal
IA	Interim Analysis	Protocol or study team required	• Helps team decide on next steps of the trial • Contains: • Safety and efficacy analysis	Per protocol or study team requirement	Critical	Either formal or informal
CSR	Clinical Study Report	FDA and protocol required	• More info: CSR • Contains: • EVERYTHING; the FULL PACKAGE • Analyses for ALL ENDPOINTS described in protocol • At a minimum, deliver reports for the primary and secondary endpoints • Exploratory endpoints are optional and whether to include depends on the type of CSR and the teams' decision	• Required when supporting other study's submission even if the study is ongoing • Each study requires a CSR as the final study deliverable, after all patients have completed/discontinued treatment and last follow-up visit is done	VERY critical	Formal

Figure 3 Keep language simple; define the acronyms

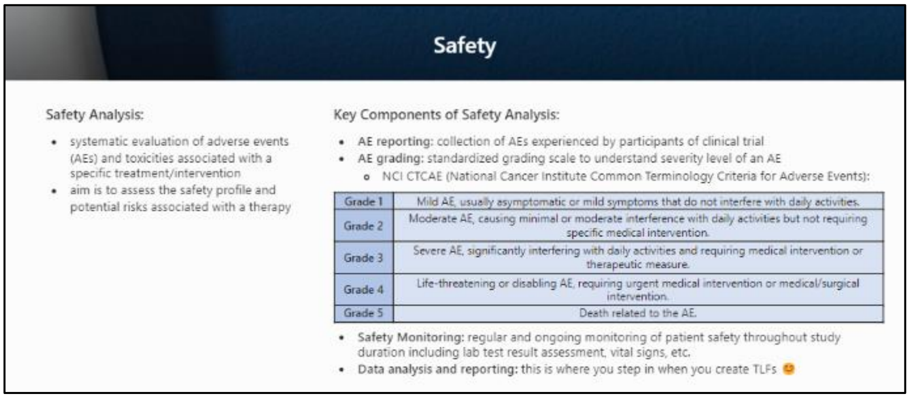
EXAMPLE #2: BLEND CONTEXT AND PRACTICALITY

Figure 4 is from a page dedicated to safety outputs, a main task for statistical programmers to work on.

It includes a brief explanation of what safety analysis is, its necessity, and some key components. The table displays common safety tables that programmers in the Oncology Early-Stage Development SP team will encounter, which macro is used to generate it, the ADaM inputs, and a description.

While a deep understanding of the medical background for each output is not required for this programmer role, having some contextual knowledge can enhance one's intuition when verifying the accuracy of the outputs. Grasping the content of the outputs can also help add interest and meaning to one's perception of their work.

In addition, the table can be highly beneficial for practical tasks in deliverable work. For example, when one updates ADAE, this table may be referenced to determine which outputs involve ADAE and therefore need to be rerun. This is especially useful when a programmer is still getting accustomed to the relationships between various ADaM datasets and their corresponding outputs.



Safety

Safety Analysis:

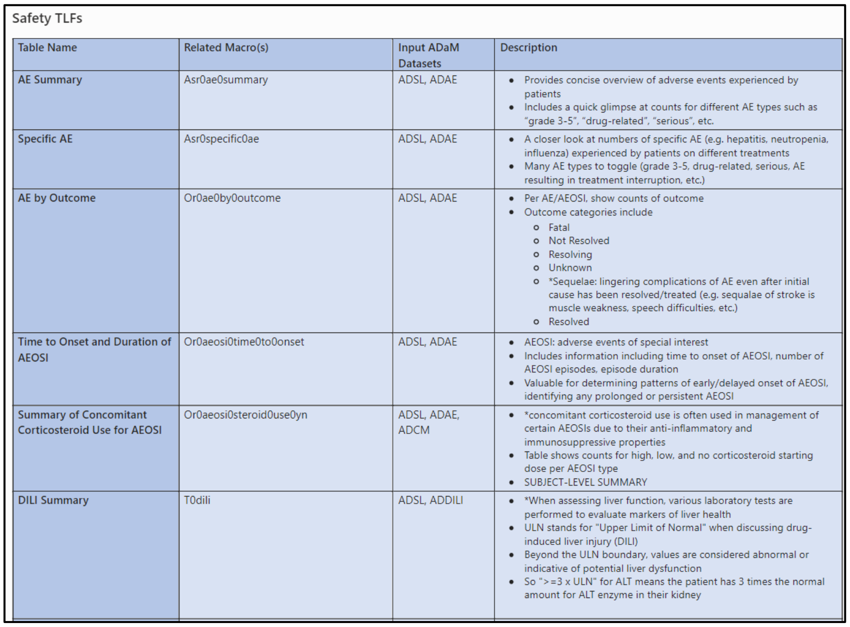
- systematic evaluation of adverse events (AEs) and toxicities associated with a specific treatment/intervention
- aim is to assess the safety profile and potential risks associated with a therapy

Key Components of Safety Analysis:

- **AE reporting:** collection of AEs experienced by participants of clinical trial
- **AE grading:** standardized grading scale to understand severity level of an AE
  - NCI CTCAE (National Cancer Institute Common Terminology Criteria for Adverse Events):

Grade 1	Mild AE, usually asymptomatic or mild symptoms that do not interfere with daily activities.
Grade 2	Moderate AE, causing minimal or moderate interference with daily activities but not requiring specific medical intervention.
Grade 3	Severe AE, significantly interfering with daily activities and requiring medical intervention or therapeutic measure.
Grade 4	Life-threatening or disabling AE, requiring urgent medical intervention or medical/surgical intervention.
Grade 5	Death related to the AE.

- **Safety Monitoring:** regular and ongoing monitoring of patient safety throughout study duration including lab test result assessment, vital signs, etc.
- **Data analysis and reporting:** this is where you step in when you create TLFs 😊



Safety TLFs

Table Name	Related Macro(s)	Input ADaM Datasets	Description
AE Summary	Asr0ae0summary	ADSL, ADAE	• Provides concise overview of adverse events experienced by patients • Includes a quick glimpse at counts for different AE types such as "grade 3-5", "drug-related", "serious", etc.
Specific AE	Asr0specific0ae	ADSL, ADAE	• A closer look at numbers of specific AE (e.g. hepatitis, neutropenia, influenza) experienced by patients on different treatments • Many AE types to toggle (grade 3-5, drug-related, serious, AE resulting in treatment interruption, etc.)
AE by Outcome	Or0ae0by0outcome	ADSL, ADAE	• Per AE/AEOSI, show counts of outcome • Outcome categories include o Fatal o Not Resolved o Resolving o Unknown o *Sequelae: lingering complications of AE even after initial cause has been resolved/treated (e.g. sequalae of stroke is muscle weakness, speech difficulties, etc.) o Resolved
Time to Onset and Duration of AEOSI	Or0aeosi0time0to0onset	ADSL, ADAE	• AEOSI: adverse events of special interest • Includes information including time to onset of AEOSI, number of AEOSI episodes, episode duration • Valuable for determining patterns of early/delayed onset of AEOSI, identifying any prolonged or persistent AEOSI
Summary of Concomitant Corticosteroid Use for AEOSI	Or0aeosi0steroid0use0yn	ADSL, ADAE, ADCM	• *concomitant corticosteroid use is often used in management of certain AEOSIs due to their anti-inflammatory and immunosuppressive properties • Table shows counts for high, low, and no corticosteroid starting dose per AEOSI type • SUBJECT-LEVEL SUMMARY
DILI Summary	T0dili	ADSL, ADDILI	• *When assessing liver function, various laboratory tests are performed to evaluate markers of liver health • ULN stands for "Upper Limit of Normal" when discussing drug-induced liver injury (DILI) • Beyond the ULN boundary, values are considered abnormal or indicative of potential liver dysfunction • So ">=3 x ULN" for ALT means the patient has 3 times the normal amount for ALT enzyme in their kidney

Figure 4 Background knowledge paves the way for meaning

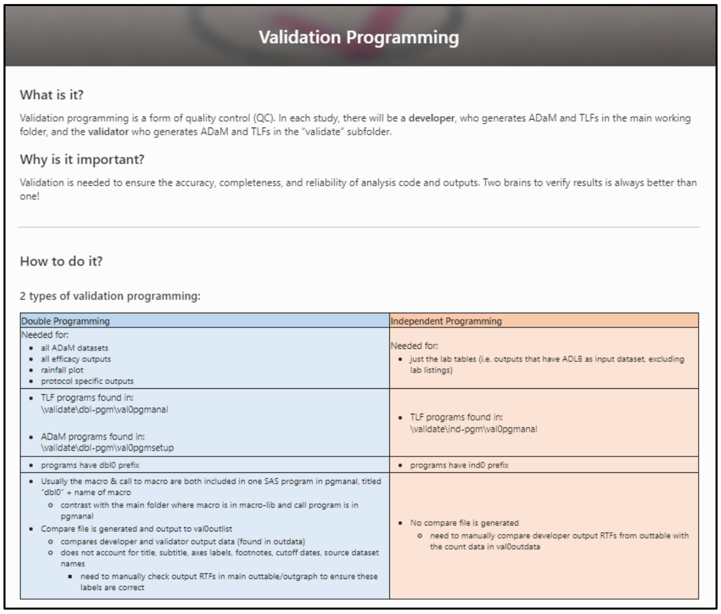
EXAMPLE #3: WHEN IN DOUBT, ANSWER THE 3 PRIMARY QUESTIONS

This next section follows 3 simple steps to introduce validation programming. Each step answers one of our primary questions.

(1) Provide a straightforward, concise definition

(2) Explain why it is necessary

(3) Provide hands-on details and instructions



Validation Programming

What is it?

Validation programming is a form of quality control (QC). In each study, there will be a **developer**, who generates ADaM and TLFs in the main working folder, and the **validator** who generates ADaM and TLFs in the "validate" subfolder.

Why is it important?

Validation is needed to ensure the accuracy, completeness, and reliability of analysis code and outputs. Two brains to verify results is always better than one!

How to do it?

2 types of validation programming:

Double Programming	Independent Programming
Needed for: • all ADaM datasets • all efficacy outputs • rainfall plot • protocol specific outputs	Needed for: • just the lab tables (i.e. outputs that have ADLB as input dataset, excluding lab listings)
• TLF programs found in: \validate\dbl-pgm\val0pgmanal • ADaM programs found in: \validate\dbl-pgm\val0pgmsetup	• TLF programs found in: \validate\ind-pgm\val0pgmanal
• programs have dbl0 prefix	• programs have ind0 prefix
• Usually the macro & call to macro are both included in one SAS program in pgmanal, titled "dbl0" + name of macro ○ contrast with the main folder where macro is in macro-lib and call program is in pgmanal • Compare file is generated and output to val0outlist ○ compares developer and validator output data (found in outdata) ○ does not account for title, subtitle, axes labels, footnotes, cutoff dates, source dataset names ▪ need to manually check output RTFs in main outtable/outgraph to ensure these labels are correct	• No compare file is generated ○ need to manually compare developer output RTFs from outtable with the count data in val0outdata

Figure 5 Answer the 3 main questions in concise language

METHODS

It is recommended to involve both senior and junior programmers in developing the handbook. The junior programmer will be able to identify pain points, and the senior programmer will know how to provide additional explanations. Over time, this material can be continually refined, as more new hires use it and give feedback on what would be helpful to include, exclude, and modify the presentation of.

The SharePoint site we created for our group was an effort of compiling all the notes I took within my first year on the job and organizing them in a presentable way. I had many questions, and the answers were not readily available in the training material offered at the time. Luckily, my mentor had many of the answers, and I recorded them in my notes, phrasing things in a way that made sense to me as a new programmer. I view this project as "creating the training material I wish I had when I first joined."

As mentioned in the SharePoint's landing page (Figure 1), please note that the website content intentionally uses simplified and accessible language for knowledge-sharing purposes only and should not be considered standard operating procedures or official documentation, as it may not cover all details or the most current technical information. Always follow your own company's SOP, as it will vary from the examples shown in this paper.

It is also important to remember that this training material does not replace company-mandatory trainings and administrative onboarding. In addition, new hires in due time become experienced hires who are ready to gain more and more cross-functional knowledge. For this purpose, it is encouraged to include a "further learning" section that points to other more in-depth resources on topics that are helpful for future learning, but not directly related to the employee's immediate everyday tasks.

CONCLUSION

Creating accessible and well-designed training materials for new hires yields substantial benefits. By prioritizing centralization, new employees will feel assured that they will not overlook essential resources. Streamlining information to focus on relevant concepts promotes more efficient and targeted learning. Focusing on foundational knowledge helps establish core concepts early, allowing new hires to recognize the significance of their work. A clearly defined pathway empowers them to work independently, reducing the number of questions posed to mentors and enabling those mentors to concentrate on their own projects. Ultimately, this comprehensive resource is invaluable for the entire team.

ACKNOWLEDGMENTS

We would like to thank our managers, Xiaohui Wang and Lily Zhang, and our colleagues, Huei-Ling Chen, Tarak Patel, Jeff Xia, and our Early-Stage Development Oncology SP team, for their support and feedback in the creation of the training website and this paper.

CONTACT INFORMATION

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

Ingrid Shu
Merck & Co., Inc., Rahway, NJ, USA
ingrid.shu@merck.com

Xinhui Zhang
Merck & Co., Inc., Rahway, NJ, USA
xinhui.zhang@merck.com