



Vlada Hanchar

**Programação de datasets de dados clínicos para
submissão regulamentar de Ensaio Clínico de
Phase I**

**Programming clinical data datasets for regulatory
submission of Phase I Clinical Trials**



Universidade de Aveiro
Ano 2023

Vlada Hanchar

**Programação de datasets de dados clínicos para
submissão regulamentar de Ensaio Clínicos de
Phase I**

**Programming clinical data datasets for regulatory
submission of Phase I Clinical Trials**

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Bioinformática Clínica ramo do Apoio à Decisão Clínica, realizada sob a orientação científica da Doutora Ana Gabriela da Silva Cavaleiro Henriques, Professora Auxiliar do Departamento de Ciências Médicas da Universidade de Aveiro

"I saw my life branching out before me like the green fig tree in the story. From the tip of every branch, like a fat purple fig, a wonderful future beckoned and winked."

Sylvia Plath

o júri

presidente

Prof. Doutor Bruno Miguel Alves Fernandes do Gago
professor auxiliar da Universidade de Aveiro

vogais

Prof. Luís Miguel Almeida da Silva
professor auxiliar da Universidade de Aveiro

Prof. Doutora Ana Gabriela da Silva Cavaleiro Henriques
professor auxiliar da Universidade de Aveiro

agradecimentos

À Universidade de Aveiro e ao Professor Bruno Gago, Diretor do Mestrado em Bioinformática Clínica, pela incansável dedicação em melhorar o rumo dos alunos.

À minha orientadora, Professora Ana Gabriela Henriques, pela disponibilidade para garantir um trabalho de qualidade.

Ao Professor Doutor Luís Almeida pela oportunidade de estagiar na BlueClinical e pela partilha da sua visão.

Ao meu orientador e chefe de departamento na BlueClinical, Bernardo Almeida, por apostar no meu progresso profissional.

À minha “chefinha”, excelentíssima Verónica Mendonça, pela paciência e dedicação nos ensinamentos e por ter feito o meu percurso uma experiência agradável.

A toda a equipa da BlueClinical e ao “melhor departamento” CDP. E em especial, à Carolina Pereira pelo companheirismo nesta nova experiência.

palavras-chave

dados clínicos, submissão regulamentar, CDISC, SDTM, ADaM, randomização, BlueClinical, Phase I, estudos clínicos, standardização de dados.

resumo

Este relatório fornece uma visão geral da experiência adquirida durante o estágio curricular na BlueClinical, como programadora de dados clínicos. Este estágio decorreu no âmbito do mestrado em Bioinformática Clínica: ramo de Apoio à Decisão Clínica, na Universidade de Aveiro. Descreve a estrutura organizacional da empresa e o papel específico do departamento de Programação de Dados Clínicos, bem como o seu enquadramento na missão geral da empresa. Além disso, o estado de arte da standardização dos dados é analisado, com foco nos requisitos regulamentares. Adicionalmente, este relatório detalha as atividades específicas realizadas ao longo do meu estágio na BlueClinical, nomeadamente a randomização dos ensaios, a criação de *datasets* e a manipulação dos dados para apresentação em relatórios. A compreensão dos requisitos regulatórios para as boas práticas clínicas, da importância da standardização dos dados e dos algoritmos de programação, permitiu a aquisição de habilidades necessárias que respondem as demandas do atual panorama de ensaios clínicos a nível mundial.

keywords

Clinical data, regulatory submission, randomization, CDISC, SDTM, ADaM, BlueClinical, Phase I, clinical trials, data standardization.

Abstract

This report provides an overview of the experiences and insights gained during my internship as a clinical data programmer trainee at Blueclinical. This internship is within the scope of the master's degree in Clinical Bioinformatics: Clinical Decision Support, in the University of Aveiro. It examines the company's organizational structure and functions, as well as the specific role of the Data Programming department within the company's broader mission. Furthermore, the state of the art of data standardisation is thoroughly reviewed, with a focus on regulatory requirements. Additionally, it highlights the key projects and tasks that I have developed during the time spent at Blueclinical, namely trial randomization, dataset generation and data manipulation for reporting. Understanding the regulatory requirements for good clinical practice, the importance of data standardization, and exploring algorithms in programming translated into the acquisition of necessary skills that respond to the demands of the current working industry of clinical trials.

Table of Contents

I. INTRODUCTION.....	1
1. Host Entity – BueClinical Ltd	1
II. STATE OF THE ART	4
1. Ethical and Regulatory Context.....	4
1.1. Development of Good Clinical Practices – Brief History.....	4
1.2. Regulatory Authorities	6
2. CDISC Development	7
2.1. SDTM Basics	8
2.2. ADAM Basics	16
III. INTERNSHIP: CLINICAL DATA PROGRAMMER ACTIVITIES.....	20
1. SAS Language	22
2. CDP Workflow	24
2.1. Randomization	26
2.2. Importing and mapping data	28
2.3. SDTM Datasets	28
2.4. ADaM Datasets	35
2.5 Tables Figures and Listings (TFLs)	40
2.6. Pinnacle validation	41
2.7. Internal Quality Control	41
IV. CONCLUDING REMARKS	42

List of Figures

Figure 1: SDTM domains per category.	12
Figure 2: Key terms of a clinical trial.	13
Figure 3: Visual representation of "JOIN" function in SQL language.	23
Figure 4: Randomization and typical workflow of clinical data programming activities. ..	25
Figure 5: Output of a blocked randomization list in a two treatment, parallel study.	27
Figure 6: Example of an output of SDTM Demographics (DM) domain dataset.	31
Figure 7: Example of an output of SDTM Exposure (EX) domain dataset.	33
Figure 8: Example of an output of SDTM Adverse Event (AE) domain dataset.	33
Figure 9: Example of an output of SDTM Pharmacokinetic Concentrations (PC) domain dataset.	34
Figure 10: Example of an output of SDTM Pharmacokinetic Parameters (PP) domain dataset.	35
Figure 11: Example of an output of ADaM Subject-Level (ADSL) dataset.	37
Figure 12: Example of an output of ADaM Plasma Concentration (ADPC) dataset.	38
Figure 13: Example of an output of ADaM Plasma Parameters (ADPP) dataset.	39
Figure 14: Example of an output of ADaM Adverse Event (ADAE) dataset.	39
Figure 15: Example of Treatment-Emergent Listing for a Clinical Study Report.	40

List of Codes

Code 1: SAS code syntax.	22
Code 2: Simplified SAS code for generating a blocked randomization list in a two-treatment (Test and Reference Product), parallel study.	27
Code 3: Simplified SAS code to generate SDTM Demographics (DM) domain dataset. ..	31
Code 4: Simplified SAS code to generate Exposure (EX) dataset.	32
Code 5: Simplified SAS code for generating ADaM Subject-Level (ADSL) dataset.	37
Code 6: Simplified SAS code for specific SAP specified rules.	38

List of Tables

Table 1: Examples of ISO-8601-compliant representations of date/time and durations....	10
Table 2: Examples of SAS PROC step and its purpose.	22
Table 3: Useful SAS functions and their purpose.	24

List of Abbreviations

ADAE	Adverse Event Analysis Dataset
ADaM	Analysis Data Model
ADaMIG	Analysis Data Model Implementation Guide
ADCM	Concomitant Medication Analysis Dataset
ADMH	Medical History Analysis Dataset
ADPC	Pharmacokinetic Concentrations Analysis Dataset
ADPP	Pharmacokinetic Parameters Analysis Dataset
ADSL	Subject-Level Analysis Dataset
ADQS	Questionnaire Analysis Dataset
AE	Adverse Event
BDS	Basic Data Structure
CDASH	Clinical Data Acquisition Standards Harmonization
CDISC	Clinical Data Interchange Standards Consortium
CDM	Clinical Data Management
CDP	Clinical Data Programming
CEIC	Comissão de Ética para a Investigação Clínica
C_{max}	Maximum concentration

CNPD	Comissão Nacional de Proteção de Dados
CO	Comments
(e)CRF	(electronic) Clinical Report Form
CRO	Clinical Research Organization
CRP	Clinical Research Partnership
CRT	Case Report Tabulation
CSP	Clinical Study Protocol
CSR	Clinical Study Report
CT	Controlled Terminology
CTIS	Clinical Trials Information System
DM	Demographics
DS	Disposition
EMA	European Medicines Agency
EU	European Union
EudraCT	European Union Drug Regulating Authorities Clinical Trials Database
EX	Exposure
FDA	Food and Drug Administration
GCP	Good Clinical Practices

ICH	International Council for Harmonisation
IND	Investigational New Drug
INFARMED	Autoridade Nacional do Medicamento e Produtos de Saúde
ISO	International Organization for Standardization
LLOQ	Lower Limit of Quantification
MedDRA	Medical Dictionary for Regulatory Activities
MW	Medical Writing
OCCDS	ADaM Structure for Occurrence Data
ODM	Operational Data Model
PC	Pharmacokinetic Concentrations
PHI	Phase I
PK	Pharmacokinetics
PP	Pharmacokinetic Parameters
PRM	Protocol Representation Model
QA	Quality Assurance
QC	Quality Control
RA	Regulatory Affairs
RELREC	Related Records

R&D	Research and Development
SAP	Statistical Analysis Plan
SAS	Statistical Analysis System
SDS	Submission Data Standards
SDTM	Study Data Tabulation Model
SDTMIG	Study Data Tabulation Model Implementation Guide
SE	Subject Elements
SEND	Standard for Exchange of Nonclinical Data
SM	Subject Disease Milestones
SOP	Standard Operating Procedure
SQL	Structured Query Language
SUPPAE	Supplementary Adverse Events Dataset
SV	Subject Visits
TA	Trial Arms
TD	Trial Disease Assessment
TDM	Trial Design Model
TE	Trial Elements
TFLs	Tables, Figures and Listings

TI	Trial Inclusion/Exclusion Criteria
TM	Trial Disease Milestones
T_{max}	Time to peak drug concentration
TS	Trial Summary
US	United States
WMA	World Medical Association
XML	Extensible Markup Language

I. INTRODUCTION

This report is the culmination of my experiences as a clinical data programmer intern within the scope of the final year of the Master's degree in Clinical Bioinformatics (Clinical Decision Support) at the University of Aveiro. During this nine-month internship period at BlueClinical, I was able to acquire experience in the working field and gain new skills for personal and professional development. My primary activities at the company allowed me to further develop my programming skills that had been a focus of my master's degree, whilst also gaining a better understanding of clinical trial procedures and the current industry requirements.

This report aims to provide a comprehensive understanding of the specific role of a clinical data programmer in the context of clinical research. It will provide a background on clinical research, as well as a discussion of the ethical and regulatory considerations involved. Additionally, it will explore the state of the art in clinical data standardization, which is a critical aspect of data submission to regulatory bodies. By examining these aspects and providing detailed description of the activities associated, this report will give insight into programming in this field.

1. Host Entity – BlueClinical Ltd

BlueClinical is a full-service contract research organization (CRO) based in Portugal. The company is prominent in the national panorama of clinical trials, specializing in providing comprehensive clinical research services to the pharmaceutical, biotechnology, and medical device industries. Founded in May 2012 and headquartered in Matosinhos, BlueClinical's aim is to support the development of innovative and effective therapies by conducting high-quality clinical trials that adhere to regulatory standards and be acknowledged as a source of smart innovation in translational medicine (“*from bench to bedside*”). BlueClinical covers all phases of the pharmaceutical research and development process, from preclinical development to commercialization, and all stages of conducting a clinical trial, from study protocol design to clinical study reports, with a special focus on early phase clinical trials.¹

To achieve its purpose, BlueClinical operates within a highly regulated environment and is subject to various guidelines and standards, including Good Clinical Practice (GCP), International Conference on Harmonisation (ICH), and regulatory requirements from the Food and Drug Administration (FDA) and the European Medicines Agency (EMA). The company's commitment to regulatory compliance ensures that all data-related activities are conducted ethically and in accordance with the highest standards of data integrity, accuracy, and confidentiality. BlueClinical is organized into three business units, each with its own mission.²

R&D – Research and Development

BlueClinical's R&D offers consultancy services and is focused on supporting institutions and companies in the research and development of new health products. The mission of the unit is to ensure successful translational medicine research and contribute to the transformation of basic research into practical health solutions, with the aim of bringing new therapeutic and diagnostic products and services to the market.³

Specific activities of BlueClinical R&D unit include providing advice on the preparation and implementation of pharmaceutical, preclinical, clinical, and regulatory development plans for new health products, as well as monitoring scientific and regulatory advice. The unit also assists in planning and supervising pharmaceutical development and analytical methods and prepares important documents such as the Investigator's Brochure and the Investigational Medicinal Product Dossier.³ These documents compile information about the investigational product in a comprehensive manner and give a better grasp of many important aspects of the protocol and reasoning to adhere to it.⁴

CRP – Clinical Research Partnerships

BlueClinical CRP provides clinical research institutions with organizational and operational support, promoting its development, efficiency, and excellence.

The CRP team consists of experienced clinical research professionals who are knowledgeable about all aspects of the clinical research process, including study design,

protocol development, regulatory compliance, site selection and management, study monitoring and data handling.⁵

Currently, there are 11 national partnerships established (with plans to expand internationally) with a cumulative experience of over 400 clinical trials in patients, across a wide range of therapeutic indications.⁵

PHI – Phase I

BlueClinical's PHI is the nucleus where conception, planning, and development of phase I clinical trials occurs. At a national level, BlueClinical has come to fill a noticeable gap in the clinical research field by investing in early-stage trials, the least conducted studies in Portugal. Phase I studies are conducted mainly on healthy subjects or selected patient populations as they are the earliest stage studies done on humans with intent to:

- (a) assess safety, tolerability and pharmacokinetics of new pharmaceuticals;
- (b) establish bioequivalence between a new drug and a reference product with the same active substance;
- (c) evaluate drug-drug interaction;
- (d) evaluate food consumption impact on drug absorption.⁶

Phase I headquarters are located in Hospital da Prelada, occupying the 3rd and, partially, the 4th floor with a total of 86 beds to accommodate volunteers for studies that take place in this unit. Phase I is fully equipped and has a modern, organised, efficient design for a smooth and controlled trial experience. After receiving approval from the relevant authorities, the National Ethics Committee for Clinical Research (Comissão de Ética para a Investigação Clínica - CEIC) and the National Committee for Data Protection (Comissão Nacional de Proteção de Dados - CNPD), all studies are carried out in accordance with international regulatory and ethical standards. In addition, over the years, along with CRP, the Phase I unit has gained an extensive repository of healthy subjects registered at the National Data Protection Committee (CNPD) and National Ethics Committee for Clinical Research (CEIC), which facilitates study planning.⁶

II. STATE OF THE ART

1. Ethical and Regulatory Context

1.1 Development of Good Clinical Practices – Brief History

In order to ensure that clinical trials are conducted in a manner that protects the rights and welfare of human subjects and produces reliable data, regulatory entities have been established. These regulatory organizations oversee the conduct of clinical trials, checking that the data they provide is correct and reliable, and examining and approving the protocols for such trials. Clinical trial regulations are intricate and frequently revised, reflecting regulatory authorities' constant efforts to ensure the compromise between the need to protect human subjects and the need to expedite the discovery of new treatments.⁷ In its core, clinical research is science that involves experimentation on humans and therefore how to conduct these studies in ethical ways has always been one of the main concerns.⁸

One of the first flagrant questioning of ethical conduct in human experimentation happened in Nazi Germany, what later came to be known as the Nuremberg Trials, in which doctors coerced concentration camp prisoners into gnarly experiments, without any type of consent. This tragedy led to the creation of the first code for human subject protection in clinical trials, the Nuremberg code of 1947 that began by stating "The voluntary consent of the human participant is absolutely essential". It emphasised, for the first time in history, that subjects must give explicit consent and that the benefits of research must outweigh the risks. Even though the Nuremberg code laid an important foundation for human rights protection in clinical trials, it was not enforced by law.⁹

Another example of unmonitored abuses occurred in Tuskegee Syphilis Study, in which African American men with syphilis were deliberately not treated so that researchers could track the natural progression of the disease. The study was conducted without informed consent, and participants were not given the option to withdraw from the study even when effective treatments became available. The study was conducted for 40 years, from 1932 to 1972, and it was only with media exposure that the unethical nature of the study was widely recognized.¹⁰

Learning from history and to prevent similar misconduct in the future, the World Medical Association (WMA) developed the Declaration of Helsinki, a statement of ethical principles

for medical research involving human subjects. Since the original 1964 version of the Declaration of Helsinki, numerous revisions have been made, due to evolving ethical norms and scientific advancements.¹¹ The Declaration of Helsinki's most current revision was approved as recently as in 2013.¹²

The adoption of the Good Clinical Practice (GCP) guideline' in the 1990s further reinforced the ethical principles outlined in the Declaration of Helsinki. These guidelines were the result of a conference, the ICH – “International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use” between regulatory authorities from Europe, America and Japan. The GCP guidelines provide a framework for the design, conduct, monitoring, recording, analysis, and reporting of clinical trials involving human subjects. This harmonisation had great benefits and facilitated the focus on innovation since it standardised and streamlined regulatory processes internationally, whilst ensuring that clinical trials are conducted ethically and with the highest standards of scientific rigor.¹³

Some key aspects of GCP are:

1. Importance of ethical principles, such as rights and safety of participants, as well as the obligation to obtain informed consent and the right to confidentiality.
2. Study design and appropriate conduct in enrolment of subjects, monitoring and reporting of adverse events.
3. Investigator qualifications and responsibilities.
4. Accurate data management and analysis.
5. Adequate monitoring and auditing of studies.
6. General quality assurance and data integrity.^{13 4}

1.2 Regulatory Authorities

The scope of clinical trials conducted by BlueClinical, or that the company helps to develop, involve both the European and the American markets, and therefore it is important to understand those specific regulatory frameworks and what entities manage them presently.

When it comes to conducting clinical trials in Portugal, the national regulatory authority is INFARMED (Instituto Nacional de Farmácia e do Medicamento) that is responsible for the authorization and supervision of the trials. INFARMED reviews clinical trial applications, including the trial protocol, to ensure that legal requirements are met and that GCP guidelines are being followed.¹⁴

In regards to ethical aspects there is an independent organisation, CEIC (Comissão de Ética para a Investigação Clínica), that gathers a board of medical professionals to review the protection of rights and welfare of participants to guarantee their safety.¹⁵

In addition, there is a specific data protection authority, CNDP (Comissão Nacional de Protecção de Dados), that supervises the use of personal data of the subjects and makes sure that their rights are being protected in accordance to the Portuguese laws and Constitution¹⁶.

If the investigational product is to be marketed in Europe, along with national authorities, the European Medicines Agency (EMA) is responsible for the scientific evaluation of medicines developed by all pharmaceutical companies. The EMA is a decentralised agency that assesses the quality, safety, and efficacy of these medicines before they can be approved for use in the EU. The agency also provides guidance on the conduct of clinical trials and monitors the safety of medicines once they are on the market.¹⁷

Clinical trials conducted in Portugal can also be submitted for approval to the US Food and Drug Administration (FDA) if the drug or treatment is intended for use in the US market. The FDA is responsible for evaluating and approving drugs for use in the US, and it sets rigorous standards for the safety and efficacy of new treatments. The FDA provides guidance on clinical trial design and data requirements to ensure that trials are conducted in a manner that will produce valid and reliable data.¹⁸

It is important to highlight that FDA has been actively involved in the development and implementation of CDISC (Clinical Data Interchange Standards Consortium) standards. In fact, the FDA has required CDISC data standards for all new drug applications since 2016

and has made CDISC-compliant data submissions mandatory for most regulatory submissions, with the exception of certain investigational new drug applications (INDs).¹⁹

2. CDISC Development

Clinical trials date back to ancient times, the first recorded instant being in the Bible in the “Book of Daniel”, in which is described the first open uncontrolled human experiment on dietary restrictions.²⁰ Since then, clinical research has had a long journey, but one thing remains consistent: data collection and analysis for hypothesis validation. Through history, data collection and analysis has had tremendous improvements through clinical monitoring, electronic data capture and clinical study databases, as well as the evolution of storing capacity and advances in programming algorithms. In the new technological era, pharmaceutical companies realised the need to implement standards for more efficient harmonisation of data accessibility and interoperability which ultimately would result in faster review cycles.²¹

The CDISC was formed in 1997 to develop global standard data models to improve the process of electronic acquisition and submission of clinical trial data to streamline medical product development. The consortium includes members from pharmaceutical companies, medical device manufacturers, regulatory authorities, and service providers. This consortium publishes standards that are recommended for clinical trial data to serve as a global interface. The primary focus of CDISC is standardising metadata, also known as “data about data” (i.e., Information on the data column’s names, labels, lengths, type.), along with controlled terminology to ensure semantic accuracy.²²

Even though this report will primarily focus on the Study Data Tabulation Model (SDTM) and the Analysis Dataset Model (ADaM), it is useful to consider the key aspects of the drug development lifecycle in order to understand how CDISC is overall organized and structured. For instance, study design, data collection, data exchange, analysis, reporting and archiving of clinical and non-clinical data.²¹ To assist each of these activities, the consortium nominated several teams to develop specific standards.

Submission data standards team (SDS) initially focused on submission of electronic clinical data associated with 10 safety domains identified by the FDA²¹. While the first iterations were well received, they had a few shortcomings, such as the need for more domains. This, later, resulted in the initial classification of domains into three primary data types (Events, Interventions, Findings) and in the development of the SDTM.²³

Analysis Dataset Model (ADaM) team with the primary goal of building on SDTM standards and developing metadata models and examples for analysis datasets. This includes documentation about the analytic results, such as statistical methods, transformations, assumptions, and derivations.²¹

The clinical data acquisition standards harmonization (CDASH) team that focused on the quality of raw data collection and consists of a set of collection formats and structures for traceable submission of data into the Study Data Tabulation Model (SDTM).²⁴

The Protocol Representation Model (PRM) team that identified, defined, and described several protocol elements such as study design and eligibility criteria.²⁵

The Controlled Terminology (CT) team which provided conventional vocabulary organised by codelists to populate the different datasets. A critical step in achieving semantic interoperability in health care.²⁶

The Operational Data Model (ODM) team that developed standards for the transfer of operational data, including administrative data and audit trail. Their efforts, later, set the stage for submission of Case Report Tabulation Data Description Specification in the Extensible Markup Language (XML) format, known as ‘define.xml’. Case Report Tabulations (CRTs) are a phrase and conceptual framework that have their roots in FDA regulation from the late 1980s. When CRTs were first developed, sponsors were supposed to be able to provide data tabulations to the FDA rather than duplicate Case Report Forms (CRFs) for each patient. The definition has recently been broadened to include data listings, patient profiles, data tabulations, and analysis datasets. The XML technology is a neutral, platform-independent, machine-readable language that allowed for easier data exchange and reading.^{21,23}

The Standard for Exchange of Nonclinical Data (SEND) team that focuses on data collected from preclinical toxicology studies and has an underlying foundation model aligned with SDTM.²⁷

2.1. SDTM Basics

CDISC Submission Data Standard for human clinical trials consists of two documents: the Study Data Tabulation Model, and the Study Data Tabulation Model Implementation Guide (SDTMIG). Over the years, the consortium has regularly released updated new versions of the standards.²⁸ The transition to a new version takes a considerable amount of time and is not required immediately upon release. This information is documented in FDA Standards Catalog with specific dates for implementation support and requirement.¹⁹ For example, during my time at the company we were just beginning to implement SDTMIG v3.3 rules for SDTM v1.7 released in November 2018, mandatory for studies beginning on 15th March of 2023, which will be my focus.

SDTM is the model that defines standard data tabulations, in which clinical information and observation from a study is compiled. SDTMIG sets a guide for implementation and helps with interpretation of SDTM's organization of raw clinical data to be submitted to regulatory authorities. The SDTMIG contains domain models and examples for frequently submitted datasets, detailed assumptions to aid with their implementation and thorough details on how to represent relationships between the datasets.²⁸ SDTMIG is built on the key concepts presented below.

Observations and Variables

Observations collected for all subjects are described by a series of variables (columns), corresponding to a row in a dataset. For each dataset, the implementation guide lists which variables are required (cannot be null), expected (even if null, must be included), or permissible (these can be conditional based on the study characteristics, or simply optional). The variable name, label and type is also specified. When applicable, a variable may specify what controlled terminology to use, indicating a codelist for it, or a format.²⁸

The key variables across all datasets are USUBJID (unique identifier for a subject), DOMAIN (two-letter dataset name abbreviation), STUDYID (usually, the protocol code), and –SEQ (a sponsor-defined numeric identifier unique within study, subject, and domain, where the hyphens are the two-letter domain code). Some common suffixes can be found in variable names, such as code variables (--CD), flag variables (--FL) and date/time variables (--DTC).²⁸

Date and/or time variables, as well as intervals and durations are required to be in ISO 8601 format [YYYY-MM-DDTHH:MM:SS]. When one of the components of a date/time variable is missing it is represented by right truncation or, in the case of intermediate components, by hyphens. Intervals are defined by two fixed instances that represent the limits of a time period. The interval of time is represented by adding a slash “/” between the two timepoints. The duration of time between the two timepoints can easily be calculated and does not need to be added to datasets, unless it is collected as a protocol defined variable. In those cases, the duration timing variable (--DUR) must be populated as [PnX], where [P] designates duration, [n] is a positive number or 0, and [X] represents any day time component (M)onth, (D)ay, (Y)ear, etc.²⁸ Some examples are depicted in Table 1.

Table 1: Examples of ISO-8601-compliant representations of date/time and durations

Date and Time/Duration collected	ISO 8601 format
May 13, 2022 09:21:18.123	2022-05-13T09:21:12.123
The 13 th of an unknown month in 2022	2022---13
Between 26 th of April and May 2 nd in 2022	2022-04-26/2022-05-2
2 Weeks 3 Days 4 Hours	P2W3DT4H

According to the information that the variable conveys, it can be classified into five major roles:

- Identifier variables, such as those that identify the study, subject, domain, and sequence number of the record.
- Topic variables, which specify the focus of the observation (such as the name of a lab test --TESTCD).
- Timing variables, which describe the timing of the observation (such as start date and end date or visit).
- Rule variables, which express an algorithm or executable method to define start, end, and branching or looping conditions in the Trial Design model.

- Qualifier variables, which include additional illustrative text or numeric values that describe the results or additional traits of the observation (such as units or descriptive adjectives). These variables can be further categorized into five sub-classes:
 - Grouping Qualifiers are used to group together a collection of observations within the same domain. Examples include --CAT and --SCAT.
 - Result Qualifiers describe the specific results associated with the topic variable (--ORRES, --STRESC, and -- STRESN).
 - Synonym Qualifiers specify an alternative name for a particular variable in an observation. (i.e. --TEST and --LOINC, which are equivalent terms for a --TESTCD).
 - Record Qualifiers define additional attributes of the observation record. Examples include Serious Adverse Events flag variables in the AE domain.
 - Variable Qualifiers are used to further modify or describe a specific variable within an observation, such as units and limit quantifications.²⁸

Domains

Groups of logically related observations are defined as domains and will often pertain to a particular SDTM dataset.²⁸ The SDTM model groups domains into distinct categories that are shown in **Figure 1**.

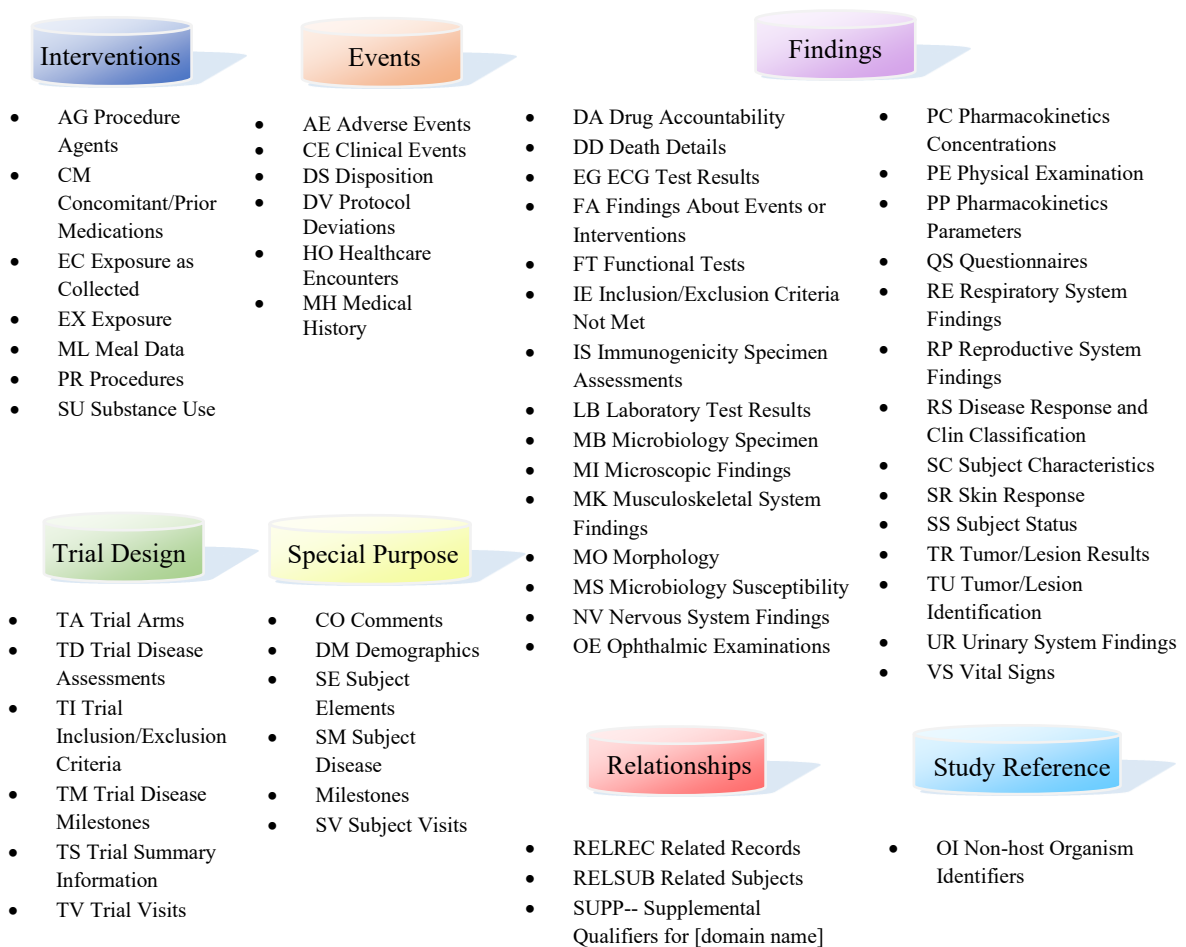


Figure 1: SDTM domains per category.

According to the implementation guide there are five main categories, based on the characteristics of the domains.

1. The general observation classes:

- **Interventions** - captures the treatments that are administered. This reflects not only exposure to the investigational product but also concomitant medications and self-administered drugs.
- **Events** - planned protocol milestones such as randomization and occurrences during (e.g.: adverse events) and prior to the trial (medical history).

- **Findings** - observations that result from planned evaluations that address specific tests or questions (lab tests, questionnaires).²⁸

2. Trial Design Model (TDM) datasets that include experimental design, schedule for assessments and trial summary and eligibility. The TDM allows for a clear, concise, and objective description of the trial in a standardised way and facilitates comparison of different trial designs. Based on the organization of conventional features into specific variables it is possible to search databases for trials with those features. Additionally, these datasets are essential in modelling planned and actual treatments and visits for subjects in a trial.²⁸

The supra mentioned datasets are built upon key terms (**Figure 2**) for any clinical trial, such as:

- **Elements** - the basic building blocks a trial has been separated into, associated with activities that affect the subjects. There are no gaps between elements as they represent the entirety of the study and, therefore, correspond to both interventional and non-interventional intervals of time.
- **Arm** - planned sequence of elements.
- **Visit** - clinical encounter that incorporated both planned and unplanned procedures and assessments.
- **Epoch** - a period of time transversal to elements and arms that apply to the trial as a whole.²⁸

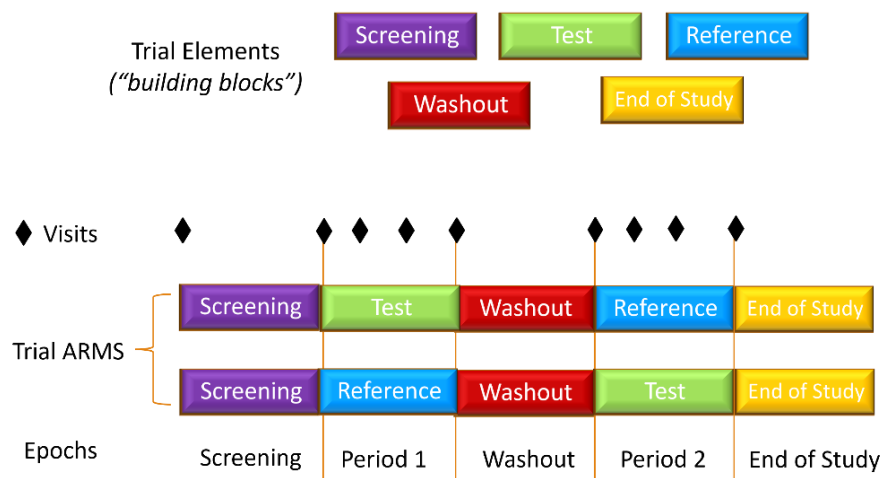


Figure 2: Key terms of a clinical trial.

The TDM datasets include:

- Trial Arms (TA) - contains each planned sequence of elements in order, also referred to as arm.
- Trial Elements (TE) - contains detailed pertinent information of all elements in a study.
- Trial Visits (TV) - visit definition and the planned order within each study arm.
- Trial Disease Assessment (TD) - information on planned disease evaluation (in accordance with protocol set timeline) and actual occurrence of disease assessments, to determine if there is good schedule compliance.
- Trial Disease Milestones (TM) - events or activities that are anticipated in the course of a disease.
- Trial Inclusion/Exclusion Criteria (TI) - contains one observation for each inclusion/exclusion criteria, not subject oriented.
- Trial Summary (TS) - one record for each characteristic that summarises the trial, not subject oriented.²⁸

3. Special Purpose Domain category includes specific data structures that do not fit into other categories:

- Comments (CO) - assembly of all comments gathered alongside other data.
- Demographics (DM) - The parent domain for all observations in relation to human clinical subjects. A set of crucial variables that describe each subject.
- Subject Elements (SE) - Actual occurrence of elements defined in Trial Elements (TE) dataset, for each subject.
- Subject Disease Milestones (SM) - Records of the actual timing of the milestones defined in Trial Disease Milestone (TM) domain, for each subject.
- Subject Visits (SV) - Actual visits recorded for each subject, that are originally planned in Trial Visits (TV) domain, or unplanned.²⁸

4. Relationship Datasets are a special purpose dataset class that use key variables STUDYID, RDOMAIN, USUBJID along with IDVAR and IDVARVAL that describe the parent's unique identifier variable and its value. Usually, the variable used for this purpose is –SEQ, a sponsor-defined numeric identifier unique within study, subject, and domain, where the hyphens are a two-letter domain code.

- Supplemental Qualifiers (SUPP--) - captures non-standard variables and their association with parent general observation classes and the Demographics domain. In addition to the key variables described in the previous paragraph, for each record includes the name of the Qualifier (QNAM), the label (QLABEL), the value (QVAL), its origin (QORIG) and evaluator (QEVAL).
- Related Records (RELREC) - collects information on the relationship between records in two (or more) datasets. A unique character identifier is attributed to each record (stored in RELID variable).²⁸

5. Study References category groups domains that are useful to establish study-specific terminology that will be used in subject data. Thus far, three situations that fit the description have been identified:

- Device Identifiers dataset that was first introduced as part of the SDTMIG for Medical Devices.
- Non-host Organisms Identifiers domain for studies that have the need to store taxonomic nomenclature of microbes or parasites that have been experimentally determined or are previously known and used as reference.
- Pharmacogenomic/Genetic Biomarker Identifiers dataset composed of groups of genetic variations of the biomarkers.²⁸

2.2 ADaM Basics

ADaM datasets are created to accommodate statistical and scientific analysis of the study results. One of the most important tenets of ADaM is traceability, which refers to the ability to understand the relation between the data and its predecessor (e.g.: SDTM), making sure that there is a clear sequence starting from data collection to data analysis. ADaM caters to most requirements for data analysis and is organised in a way to facilitate the presentation of results in Tables, Figures and Listings (TFLs) that incorporate the final report (CSR). For that purpose, ADaM does not have many required variables, and instead many of the standard variables are permissible, which offers more structural flexibility than SDTM.²⁹

General Variable Conventions

1. ADaM variables are standardized in the implementation guide and when directly derived from SDTM should maintain their attributes (name, label, length, etc.).
2. Character variables will often be mapped 1:1 to a corresponding numeric variable that will be named as presented in SDTM, with an added “N” suffix. This can be useful for ordering or organizing information.
3. Indicator variables (--FL) have at most two possible non-missing values, Y or N (i.e., yes or no) and may have a corresponding numeric variable (--FN). Record-level and parameter-level flags are an exception as only Y (yes) or null values are allowed.
4. The A- prefix indicates a variable related to analysis observations (e.g.: AVISIT, AVAL).
5. There is no required format for timing variables.²⁹

All ADaM datasets are named ADxx, where ‘xx’ is often carried over from the source SDTM domain, or sponsor defined. According to CDISC, ADaM defines dataset and metadata standards for efficient generation, replication, and review of clinical trial statistical analyses. Similar to the case of SDTM, there is an ADaM Implementation guide (ADAMIG v1.1) that lays out the rules for the execution of two standard data structures: the Subject-Level Analysis Dataset (ADSL) and the Basic Data Structure (BDS). The third ADaM

standard data structure, the Occurrence Data Structure (OCCDS), is described in the document ADaM Structure for Occurrence Data (OCCDS) v1.0.²⁹

Subject-Level Analysis Dataset (ADSL)

The ADSL dataset and related metadata are required for CDISC based submission even in the event of no other ADaM dataset submission. The Subject-Level Analysis is presented with one record per subject, regardless of the trial design type. This structure allows for easy merging with other datasets, such as other analysis datasets. ADSL variables contain important information in relation to subject disposition, demographic and baseline characteristics that describe a subject's experience in the trial and enable and facilitate analysis: identifier variables, subject demographics variables, population indicator variables (flags --FL), treatment variables, dose variables, treatment timing variables (including Subject-level Period, Subperiod and Phase Timing Variables) and Subject-Level Trial Experience Variables.²⁹

The required variables are: STUDYID, USUBJID, SUBJID, SITEID, AGE, AGEU, SEX, and RACE, all of which can be taken directly from the SDTM Demographics (DM) dataset. The description of the planned arm (ARM) alongside with the planned treatment for a respective period (i.e. TRT01P) is also required.²⁹

A crucial purpose of ADSL is to provide analysis population inclusion flags that are previously defined in the statistical analysis plan (SAP). At least one population flag is required. Some of the standard flags are: ENRFL (enrolled population), RANDFL (randomized population), SAFFL (Safety population), FASFL (Full Analysis Set population), PPROTFL (Per-Protocol population), and COMPLFL (Completers population).²⁹

Basic Data Structure (BDS)

The BDS dataset contains the actual analysis for the study and is structured as one or more records per subject, analysis parameter or timepoint. This structure contains parametric and nonparametric analysis and is especially useful for continuous analyses of values such as mean, median, standard deviation and other statistics. Typically, there are several BDS

datasets in a study and ADaMIG defines the standard variables that apply.²⁹ Some examples, from SDTM findings class as predecessor include ADPC (Analysis Data Pharmacokinetic Concentrations), ADPP (Analysis Data Pharmacokinetic Parameters) and ADQS (Analysis Data Questionnaires Dataset).

It is important to note some key aspects when creating a BDS dataset. First, some variables are taken directly from the SDTM dataset and others are derived. PARAM and PARAMCD variables serve a similar purpose to SDTM variables –TEST and –TESTCD, however they are created for a parameter that meets an analysis need, and not just because it is collected. Additionally, in contrast to SDTM, no further qualifiers are needed to describe PARAM, as it will already contain the information relevant for that analysis. AVAL and AVALC contain the numeric and character values, respectively, described in PARAM. In the case of no derivations or imputations needed, these values are equivalent to the ones present in SDTM variables –STRESN and –STRESC. However, that although both SDTM variables are populated does not imply that both ADaM variables should be populated as well. For most parameters, only either the numeric or the character variable is required, according to the analysis needs. In some cases, it might be useful to represent both variables when there is a one-to-one mapping, such as in the case of ADQS, where a numeric score might correspond to a questionnaire answer. The dates associated with AVAL and AVALC will have the prefix “A” and will usually be derived from SDTM timing variables. The same principle applies to the variables AVISIT and AVISITN that represent the analysis visit of the record, and will usually map to SDTM –VISIT, although not always.²⁹

Occurrence Data Structure (OCCDS)

OCCDS is designed to be analysis-ready for the typical occurrence analysis, where subjects with a given record are counted. These datasets are derived from SDTM intervention and events class datasets and usually do not fit into Basic Data Structure (BDS). According to Analysis Data Model (ADaM) Structure for Occurrence Data (OCCDS) v1.0 this is due to the following reasons:

- AVAL and AVALC are obsolete, since occurrences are counted in analysis. There are one or more records for each occurrence assessment;

- Occurrence codings are well defined in a dictionary (i.e. MedDRA) which usually implies the creation of specific well-structured categories of variables with specific terminology.
- There is no need for analysis versions of the dictionary hierarchy variables.²⁹

Examples of data that fit into this structure include those used for typical analysis of Adverse Events (ADAE), Concomitant Medications (ADCM), and Medical History (ADMH).

III. INTERNSHIP: CLINICAL DATA PROGRAMMER ACTIVITIES

During my internship at BlueClinical, I had the opportunity to work in the Clinical Data Programming & Biostatistics department (CDP), which is responsible for providing data management, programming, and statistical analysis of vast amounts of data generated during clinical trials. The department works closely and in synergy with other departments, including Pharmacokinetics (PK), Clinical Data Management (CDM), Medical Writing (MW) and Regulatory Affairs (RA), to prepare a Clinical Study Report (CSR) for submission to the regulatory authorities and to ensure that all data-related actions are carried out in accordance with legal standards throughout the clinical trial process.

The internship allowed me to acquire knowledge and develop competences in programming and gain an understanding of clinical data handling requirements. The activities carried out during the 9-month internship period reflected the experience of every new collaborator of the CDP department. BlueClinical's policy ensures that all personnel is qualified and equipped with up-to-date training. The training plan included self-study and training sessions conducted by both the company and external entities. This encompasses both generic and role-specific training.

Generic Training

The generic training serves as a baseline for familiarization with the company's Standard Operating Procedures (SOPs) and to obtain a transversal knowledge of clinical research.

SOPs are documents that provide detailed instructions and guidelines for all internal procedures within the company. Their purpose is to ensure consistency, uniformity, and quality across different activities. All employees are required to be well-versed in the general company procedures and those related to their particular role. Moreover, BlueClinical requires Quality Assurance internal training sessions that cover audit procedures, inspections, and the implementation of preventive and corrective actions when misconduct or break of confidentiality are identified.

One of the other fundamental training imperatives is a 3-hour GCP refresher course ("*Good Clinical Practice: A Refresher Course for all Site Personnel Working on Clinical Research Studies (including ICH E6, Revision 2 changes)*"), available on an external

platform, Syneos Health. This course is a comprehensive guide for all clinical personnel and depicts the basic principles of GCP discussed in the introductory chapter.

Role specific training

The specific training varies according to each department and each role. The existence of well-established norms obliges data handling to follow rigorous standards. Programming emerges as a way of optimizing the organization and presentation of clinical data, for subsequent submission to regulatory authorities. In this context, the CDP training included:

- gaining proficiency in data standard comprehension;
- programming courses certifications;
- trial randomization;
- data integration, importing and mapping;
- SDTM and ADaM datasets generation;
- TFLs creation;
- quality control of CDP activities.

As aforementioned, CDP functions are closely tied with data manipulation and know-how of various programming languages, including, but not limited to, R, Python and with a special emphasis on SAS (Statistical Analysis System). In this sense, I had the opportunity to obtain a SAS certification through an introductory course (“*SAS Programming 1: Essentials*”), one of the first required training. Simultaneously, other activities were introduced in regard to the CDP involvement in clinical trials through continuous training sessions and performance of quality control on work done by other team members.

CDP’s workflow and activities as well as the SAS language will be explored in the following sections. The first section is a SAS language crash course and provides general useful syntax and functions. The second section refers to the CDP workflow and is divided into the specific activities inherent to the job, that I had the opportunity to develop for several studies. It will also provide practical examples that illustrate the application of CDISC standards and regulatory requirements discussed previously. To preserve confidentiality, the datasets and code for generating the datasets provided are simplified and the data is simulated.

1. SAS Language

SAS (Statistical Analysis System) is a programming language widely used in clinical research and data analysis that provides a range of tools and features that make it easy for programmers to handle large datasets, perform statistical analyses, and generate reports. Furthermore, SAS has a well-established reputation in the pharmaceutical industry, and its use is highly recommended by regulatory agencies like the FDA for the submission of clinical trial data³⁰. SAS programming is based on two building blocks: DATA Step and PROC Step.

The **DATA step** can read an existing SAS dataset (SET statement) or create a dataset by defining, formatting and labelling variables (INPUT statement) and/or reading input files (INFILE statement).

```
DATA <new_data>;  
  SET <old_data>;  
  INFILE <file-specification>;  
  INPUT <variable> <variable_type> <column(s)>;  
RUN;
```

Code 1: SAS code syntax.

The **PROC step** can perform specific procedures over the data, detailed in **Table 2**.

Table 2: Examples of SAS PROC step and its purpose.

PROC step	Purpose
PROC CONTENTS	to view the general structure of a file (variables, type, length, labels).
PROC PRINT	to partially print datasets portions.
PROC SORT	to rearrange observations in datasets according to a specific order.
PROC FORMAT	to store inputted synonyms to transform data into.
PROC FREQ	to output a frequency table of categorical data.
PROC MEANS	to summarize statistical continuous data (mean, median, min, max, etc).
PROC REPORT	to create detailed reports that can be configured by the user (combines data step, statistical procedures, and printing).

PROC SQL is a particular procedure statement within SAS that calls the SQL (Structured Query Language) functions and language inside a SAS program. SQL is widely known for efficient database creation and maintenance as it allows to create and alter tables in a structured query with an intuitive syntax. The basic SQL syntax consists of:

SELECT * FROM - which specifies the columns(/variables) from a dataset. Asterisk means all columns are being selected.

WHERE - subsets the data based on a condition. The condition will often be determined by an equality '=' or statements 'IN ()', 'NOT IN ()', 'LIKE %%'.

GROUP BY - groups observations by a specific column(s).

ORDER BY - sorts observations by a specific column(s).

JOIN statements (**Figure 3**): merge datasets through key variables that datasets share.

INNER JOIN - results in observations that are common in both datasets.

OUTER JOIN - results in the merging of all observations from both datasets.

LEFT JOIN - results in all observations of the dataset defined in FROM statement and all common observations between the two datasets.

RIGHT JOIN - results in all observations of the dataset that is being merged in this statement and common observations between the two datasets.

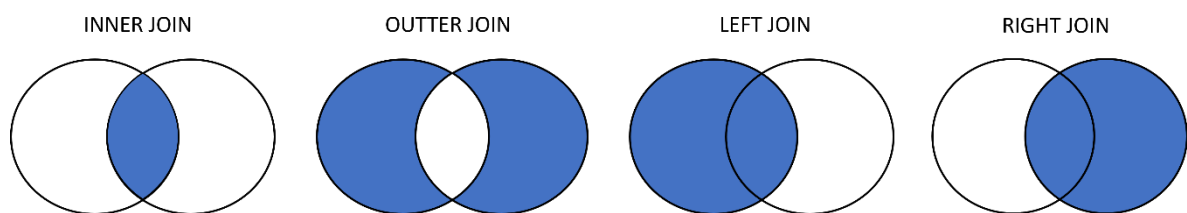


Figure 3: Visual representation of "JOIN" function in SQL language.

Additionally, SAS has a powerful programming feature referred to as Macros that is useful to avoid repetitive coding. Macro variables are used to store a value, either globally and can be called at any point in the program, or locally and can be called when declared in the program. However, a Macro also refers to a group of SAS statements with specific syntax, along with DATA and PROC steps, that define reusable code fragments that can

have a different outcome based on the parameters introduced. In this sense, the macro language is a tool for code optimization and maintenance. Other useful SAS functions are depicted in **Table 3**.

Table 3: Useful SAS functions and their purpose.

SAS Function	Purpose
PUT/INPUT (source, format)	transforms the variable from numeric to character and vice versa.
SUBSTR (argument, position, length)	returns a portion of the argument, as specified.
CATS (variable1,...,variableN)	concatenates specified variables.
UPCASE/LOWCASE (string)	formats text to capital or lower case letters, respectively.
TRANWRD (source, target, replacement)	replaces all occurrences of target expression.
TRIM (argument)	removes trailing elements from argument.

2. CDP Workflow

The CDP department at BlueClinical operates at various points of the trial pipeline (**Figure 4**), having a crucial role in how the data is reported to regulatory entities and thus having to be knowledgeable not only in regard to data science procedures but also to specific standards (CDISC) and guidelines inherent to clinical research. This mandatory approach increases quality, provides transparency and traceability whilst also diminishing errors.

As a member of the team, I was involved in various activities such as data retrieval and integration, trial randomization list generation, data mapping from various sources, dataset creation and validation, and data reporting. Additionally, quality control was performed at every task, as it is transversal to all BlueClinical functions.

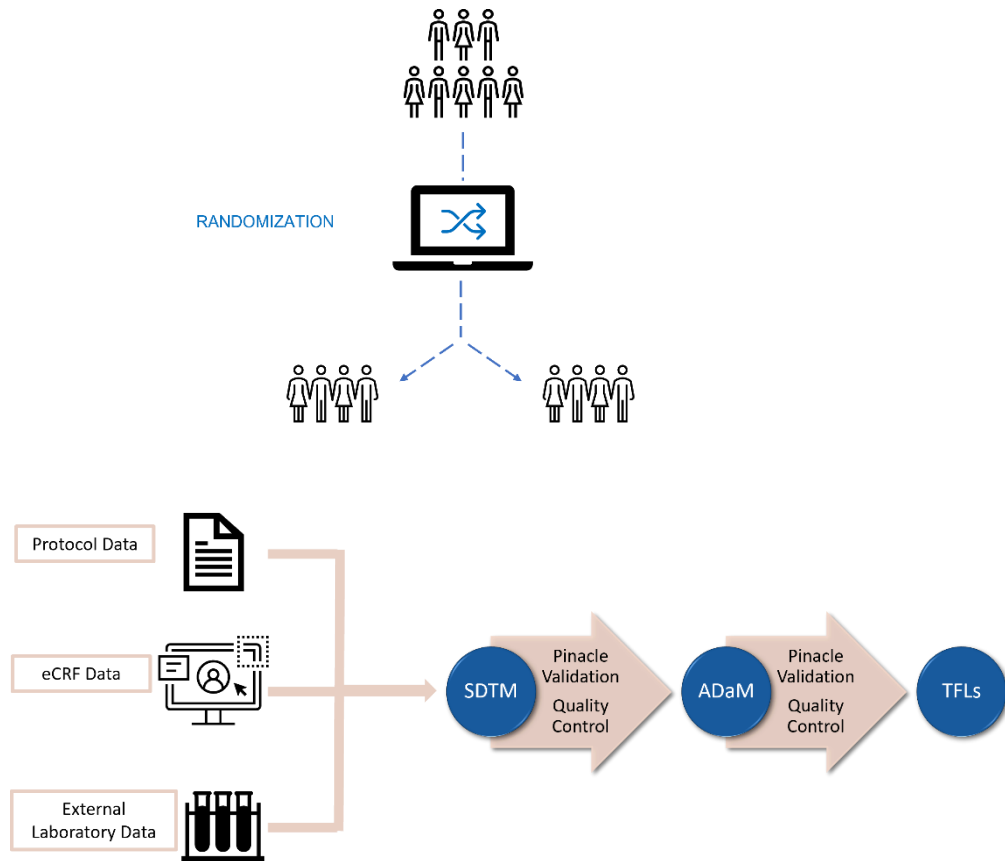


Figure 4: Randomization and typical workflow of clinical data programming activities.

To gain practical insights into the workflow of clinical data management, while upholding patient confidentiality, mock data was utilized to construct datasets. The simulated data emulates a clinical phase I trial, developed to evaluate the bioequivalence and bioavailability of a Test Product against a Reference Product already available in the market. The study design follows a parallel one-period assessment, where 12 healthy participants were randomly assigned to receive either the test or the reference product.

2.1 Randomization

A good clinical trial is one that will provide unbiased evaluations and minimise covariate variability. Randomization is a critical component of clinical trials that has become a standard practice to mitigate accidental bias. The randomization process ensures that participants are assigned to different study groups at random and that there is no systemic allocation. In practice, randomization is achieved through generating randomization schedules which must be reproducible. According to study needs, there are different types of randomizations: simple, stratified, or adaptive. Simple randomization assigns subjects to a treatment group, purely by chance, similar to flipping a coin. Stratified randomization groups subject by *stratas* (e.g., age, gender, body mass index) and allocation to treatment is performed within each subgroup.³¹ Adaptive randomization can be based on specific covariates and previous assignment of participants or on trial response.³²

A simple or stratified randomization process might, by probability, assign a different number of participants to each study group. Block randomization is a technique used to divide participants into blocks so that an equal number of participants inside each block are allocated to each treatment. Blocks are small and balanced and should be a multiple of the number of study groups or treatments. When possible, the block size is, at least, twice the number of groups or treatments.

A randomization plan is a signed document that guarantees transparency and reliability through the randomization process. This document is created based on CSP instructions and compiles the randomization parameters and steps necessary for randomization.

First, a seed must be generated and stored so that the output can be reproduced. In succession, the seed is called inside a PROC PLAN procedure. PROC PLAN is used for permutations and combinations of numbers and allows to define factors, such as the number of blocks, block size, the values that each block contains and if value attribution is random or ordered. Only when this randomization plan specification has been approved and signed, can we proceed to generating a randomization list, that must also be approved before the final signed version.

For a parallel study, in which, there are 12 patients, two treatments and an allocation ratio of 1:1, there would be 3 blocks of size 4, twice the number of treatments ($12/4=3$). A simplified code is shown in **Code 2** and the generated output in **Figure 5**.

```
/*Generating seed*/
```

```

seed=round(ranuni(0)*1000000);
call symput ('seed', seed);
run;

/*blocked randomization*/
proc plan seed=&seed.;
    factors block=3 random treatment= 4 random;
    output out=randplan
        treatment nvals=(1 1 2 2);
run;

/*Patient ID attribution and Treatment formatting*/
data randomization(keep=patientid block treatment);
    set randplan;
    PATIENTID = put(_N_, z2.);
run;

proc format;
    value $treatment
        1='Test Product'
        2='Reference Product';
run;

/*Final list output*/
proc report data=randomization;
    columns patientid
        treatment;

    define patientid / order "Patient ID";
    define treatment / display "Treatment" format=$treatment.;
run;

```

Code 2: Simplified SAS code for generating a blocked randomization list in a two-treatment, parallel study.

Patient ID	Treatment
01	Reference Product
02	Test Product
03	Test Product
04	Reference Product
05	Test Product
06	Reference Product
07	Test Product
08	Reference Product
09	Test Product
10	Test Product
11	Reference Product
12	Reference Product

Figure 5: Output of a blocked randomization list in a two treatment, parallel study.

2.2. Importing and mapping data

Once patients are randomised and the study takes course, a vast amount of data is collected, as defined by CSP. The first step to handling data is to import it into the current SAS project and start mapping variables that will later be useful for dataset merging and data interpretation. The data that is imported can be originated from CSP, eCRF, biolaboratory results of analyte concentrations, pharmacokinetic data as parameters provided by the PK team and other external sources, if defined in CSP.

The creation of libraries inside the SAS project are a useful step to store imported and mapped values and, for this purpose the respective libraries were named “raw” and “map”. Imported protocol data was stored in library “info”. An additional, temporary library for intermediate and draft datasets was named “work”.

2.3. SDTM Datasets

As established previously, SDTM datasets are a structured and standardised way of presenting raw data from a study. For each study, different data may be collected and, therefore, there will be a need to create different datasets, in accordance with what is stipulated in SDTMIG. Since there are numerous domains, as discussed previously, below are represented only some of the datasets that I generated and that are usually staples in any study (Demographics (DM), Exposure (EX), Adverse Events (AE), Pharmacokinetic Concentrations (PC) and Pharmacokinetic Parameters (PP)). In this sense, there is a natural logic to creating these datasets. The first ones to be created tend to be the TDM Datasets since they contain information directly obtained from CSP definitions and are not dependent on subject observations. Likewise, it is important to establish the DM dataset, that contains the demographic information about the subjects.

Demographics (DM)

Demographics (DM) is a special-purpose domain that describes each subject in a clinical study. This dataset will contain information about subject demographic data such as age and sex, planned and actual arms, and reference timepoints useful in other domains.

Code 3 provides a simplified query for the creation of this dataset (**Figure 6**) and includes all required and conditional variables. The query is separated into two major categories: screening failures, and subjects of differing arms. The major difference being that, although screening failures must populate DM, they won't have any timing record qualifiers, except for RFPENDTC that refers to the date/time of the end of their participation in the study. Some aspects to note in this dataset are the differences between the various – DTC variables. In order to accommodate different reference period definitions, there are three sets of reference variables. In some cases, the values of the variables— RFSTDTC and RFENDTC, RFXSTDTC and RFXENDTC- may be identical. However, RFXSTDTC and RFXENDTC always represent the first and last study exposure (calculated in 'work._exminmax', for each subject), while RFSTDTC and RFENDTC can be more flexible and represent reference date time for protocol specified procedures. In this example, RFENDTC is considered as the day and time of End of Study procedure derived from raw CRF disposition data.

One particular change from SDTMIG v3.2 to v3.3 is how the arm variables should be populated. In the current version, when ARM-- and/or ACTARM-- variables are NULL, ARMNRS should contain the respective reason. In the case of an unplanned treatment being administered, ACTARMUD must be completed with the description of that treatment.

```
proc sql;
/* Calculates first and last exposure to formulation data/time */
create table work._exminmax as
select umap.usubjid,
       min(ex.EXSTDTC) as firstex,
       max(ex.EXSTDTC) as lastex
from traw.ex as ex
inner join map.usubjmap as umap on umap.subjectid=ex.subjectid
group by ex.subjectid;

/* Selects subjects that screening failures */
create table TSDTM.DM as
select
  'BLCL-XXXX' as STUDYID,
  'DM' as DOMAIN,
  umap.USUBJID,
```

```

dm.subjectid as SUBJID,
" as RFSTDTC,
" as RFENDTC,
" as RFXSTDTC,
" as RFXENDTC,
dm.ICFDAT, " ", "T" as RFICDTC,
case when ds.dsentsim ~=" then catx('T', ds.dsendat, ds.dsentsim)
                                else ds.dsendat end as RFPENDTC,
catx('T', ds.dthdat, ds.dthtim) as DTHDTC,
dm.sitename as SITEID,
dm.age as AGE,
case when AGE~=" then 'YEARS' else " end as AGEU,
case when dm.sex='Male' then 'M' else 'F' end as SEX,
upcase(dm.race) as RACE,
" as ARMCD,
" as ARM,
" as ACTARMCD,
" as ACTARM,
'SCREEN FAILURE' as ARMNRS,
" as ACTARMUD,
'PRT' as COUNTRY
from traw.dm as dm
inner join map.usubjmap as umap on (umap.subjectid=dm.subjectid and umap.rmrandn=")
inner join raw.ds as ds on (ds.subjectid=umap.subjectid and ds.dsreas='Trial screen failure')

```

/* Subjects according to randomization to arm */

```

UNION ALL
Select 'BLCL-XXXX' as STUDYID,
'DM' as DOMAIN,
umap.USUBJID,
dm.subjectid as SUBJID,
tranwrđ(ex.exstdtc, ' ', 'T') as RFSTDTC,
case when ds.dsentsim ~=" then catx('T', ds.dsendat, ds.dsentsim)
                                else ds.dsendat end as RFENDTC,
tranwrđ(exdtc.firstex, ' ', 'T') as RFXSTDTC,
tranwrđ(exdtc.lastex, ' ', 'T') as RFXENDTC,
dm.ICFDAT, " ", "T" as RFICDTC,
case when ds.dsentsim ~=" then catx('T', ds.dsendat, ds.dsentsim)
                                else ds.dsendat end as RFPENDTC,
catx('T', ds.dthdat, ds.dthtim) as DTHDTC,
dm.sitename as SITEID,
dm.age as AGE,
case when AGE~=" then 'YEARS' else " end as AGEU,
case when dm.sex='Male' then 'M' else 'F' end as SEX,
upcase(dm.race) as RACE,
rndseq.armcd as ARMCD,
rndseq.arm as ARM,
case when rmarm.armcd~= rmarm.actarmcd then "
                                else rmarm.actarmcd end as ACTARMCD,
case when rmarm.armcd~= rmarm.actarmcd then "
                                else rmarm.actarm end as ACTARM,
case when rmarm.armcd=" then "NOT ASSIGNED"
when rmarm.armcd ~= rmarm.actarmcd then
"UNPLANNED TREATMENT" end as ARMNRS,
case when rmarm.arm~= rmarm.actarm then rmarm.actarm end as ACTARMUD,
'PRT' as COUNTRY
from raw.dm as dm
inner join map.usubjmap as umap on (umap.subjectid=dm.subjectid)

```

```

inner join map.randarm as rmarm on (rmarm.subjectid=dm.subjectid)
left join raw.ex as ex on (ex.subjectid=umap.subjectid)
left join raw.ds as ds on (ds.subjectid=umap.subjectid)
left join work._exminmax as exdtc on (exdtc.usubjid=umap.usubjid);

```

run;

Code 3: Simplified SAS code to generate SDTM Demographics (DM) domain dataset.

STUDYID	DOMAIN	USUBJID	SUBJID	RFSTDTC	RFENDTC	RFXSTDTC	RFXENDTC	RFICDTC	RFPENDTC	SITEID
BLCL-XXXX	DM	SCR_0035	0035	2022-11-19T06:30	2022-12-22T06:36	2022-11-19T06:30	2022-11-19T06:30	2022-11-02	2022-10-28T15:24	BlueClinical Phase I
BLCL-XXXX	DM	BLCL-XXXX_01	0012	2022-11-19T06:30	2022-12-22T10:34	2022-11-19T06:30	2022-11-19T06:30	2022-10-28	2022-12-22T06:36	BlueClinical Phase I
BLCL-XXXX	DM	BLCL-XXXX_02	0084	2022-11-19T06:32	2022-12-22T10:34	2022-11-19T06:32	2022-11-19T06:32	2022-10-31	2022-12-22T10:34	BlueClinical Phase I
BLCL-XXXX	DM	BLCL-XXXX_03	0015	2022-11-19T06:34	2022-12-22T06:40	2022-11-19T06:34	2022-11-19T06:34	2022-11-02	2022-12-22T06:40	BlueClinical Phase I
BLCL-XXXX	DM	BLCL-XXXX_04	0065	2022-11-19T06:36	2022-11-22T06:47	2022-11-19T06:36	2022-11-19T06:36	2022-11-02	2022-11-22T06:47	BlueClinical Phase I
BLCL-XXXX	DM	BLCL-XXXX_05	0086	2022-11-19T06:38	2022-12-22T06:44	2022-11-19T06:38	2022-11-19T06:38	2022-11-02	2022-12-22T06:44	BlueClinical Phase I
BLCL-XXXX	DM	BLCL-XXXX_06	0022	2022-11-19T06:40	2022-11-22T07:12	2022-11-19T06:40	2022-11-19T06:40	2022-10-31	2022-11-22T07:12	BlueClinical Phase I
BLCL-XXXX	DM	BLCL-XXXX_07	0040	2022-11-19T06:42	2022-11-22T06:48	2022-11-19T06:42	2022-11-19T06:42	2022-11-02	2022-11-22T06:48	BlueClinical Phase I
BLCL-XXXX	DM	BLCL-XXXX_08	0048	2022-11-19T06:44	2022-12-22T07:50	2022-11-19T06:44	2022-11-19T06:44	2022-11-03	2022-12-22T07:50	BlueClinical Phase I
BLCL-XXXX	DM	BLCL-XXXX_09	0021	2022-11-19T06:46	2022-11-22T06:52	2022-11-19T06:46	2022-11-19T06:46	2022-11-03	2022-11-22T06:52	BlueClinical Phase I
BLCL-XXXX	DM	BLCL-XXXX_10	0089	2022-11-19T06:48	2022-12-22T07:03	2022-11-19T06:48	2022-11-19T06:48	2022-11-03	2022-12-22T07:03	BlueClinical Phase I
BLCL-XXXX	DM	BLCL-XXXX_11	0090	2022-11-19T06:50	2022-11-22T06:56	2022-11-19T06:50	2022-11-19T06:50	2022-11-03	2022-11-22T06:56	BlueClinical Phase I
BLCL-XXXX	DM	BLCL-XXXX_12	0006	2022-11-19T06:52	2022-11-22T06:59	2022-11-19T06:52	2022-11-19T06:52	2022-11-03	2022-11-22T06:59	BlueClinical Phase I

USUBJID	AGE	AGEU	SEX	RACE	ARMCD	ARM	ACTARMCD	ACTARM	ARMNRS	ACTARMUD	COUNTRY
SCR_0035	29	YEARS	M	BLACK OR AFRICAN AMERICAN					SCREEN FAILURE		PRT
BLCL-XXXX_01	28	YEARS	F	WHITE	R	Reference Product	R	Reference Product			PRT
BLCL-XXXX_02	38	YEARS	F	BLACK OR AFRICAN AMERICAN	T	Test Product	T	Test Product			PRT
BLCL-XXXX_03	25	YEARS	M	WHITE	T	Test Product	T	Test Product			PRT
BLCL-XXXX_04	27	YEARS	F	WHITE	R	Reference Product	R	Reference Product			PRT
BLCL-XXXX_05	36	YEARS	M	WHITE	T	Test Product	T	Test Product			PRT
BLCL-XXXX_06	45	YEARS	M	WHITE	R	Reference Product	R	Reference Product			PRT
BLCL-XXXX_07	30	YEARS	M	WHITE	T	Test Product	T	Test Product			PRT
BLCL-XXXX_08	21	YEARS	M	WHITE	R	Reference Product	R	Reference Product			PRT
BLCL-XXXX_09	22	YEARS	F	BLACK OR AFRICAN AMERICAN	T	Test Product	T	Test Product			PRT
BLCL-XXXX_10	31	YEARS	F	WHITE	T	Test Product	T	Test Product			PRT
BLCL-XXXX_11	31	YEARS	M	WHITE	R	Reference Product	R	Reference Product			PRT
BLCL-XXXX_12	28	YEARS	F	WHITE	R	Reference Product			UNPLANNED TREATMENT	Test Product	PRT

Figure 6: Example of an output of SDTM Demographics (DM) domain dataset.

Exposure (EX)

Exposure (EX) is an intervention domain with all records pertaining to a protocol specified study treatment. **Figure 7** is an example of what an EX dataset output must look like and contains all required and expected variables. As seen in **Code 4**, EX dataset is composed of information about the treatment derived from CSP (“*info.trts*”) and only contains observations from raw data (“*raw.ex*”) where the subject has been exposed to study treatment (where “*ex.exyn*='Y'”). Unlike the other variables present in this domain, the variables EXSTDY, EXENDY are calculated outside of raw data collection. The study start/end days of treatment are incremented by 1 for each day starting from the subject reference start date (“*dm.RFSTDTC*”), with the assumption that there is no Study Day 0

(*%macro*STDY). EXSEQ is also calculated separately, since it is a unique sequence number for an observation within each subject.

```
%macro STDY(ref=, date=);  
    case when scan(&date, 1, 'T') >= scan(&ref, 1, 'T') then input(substr(scan(&ref, 1, 'T'),9,2), best.) -  
input(substr(scan(&date, 1, 'T'),9,2), best.) +1  
    else input(substr(scan(&ref, 1, 'T'),9,2), best.) - input(substr(scan(&date, 1, 'T'),9,2), best.) end  
%mend STDY;
```

```
proc sql;  
    create table WORK._EX as  
    select 'BLCL-XXXX' as STUDYID,  
        'EX' as DOMAIN,  
        umap.usubjid,  
        " as EXSEQ,  
        ex.extrt as EXREFID,  
        trts.trtprod as EXTRT,  
        trts.trtdose as EXDOSE,  
        trts.trtdoseu as EXDOSU,  
        trts.trtdosfrm as EXDOSFRM,  
        trts.trtdosfreq as EXDOSFRQ,  
        trts.trtroute as EXROUTE,  
        tranwrđ(ex.exstdtc, " ", "T") as EXSTDTC,  
        tranwrđ(ex.exstdtc, " ", "T") as EXENDTC,  
        %STDY(ref=dm.rfstdtc, date=ex.exstdtc) as EXSTDY,  
        %STDY (ref=dm.rfstdtc, date=ex.exendtc) as EXENDY,  
        'TREATMENT' as epoch  
    from RAW.EX as EX  
    inner join MAP.USUBJMAP as umap on (umap.subjectid=ex.subjectid)  
    inner join INFO.TRTS as trts on (trts.trtcode=ex.extrt)  
    left join tsdtm.dm as dm on dm.subjid=ex.subjectid  
    where ex.exyn='Yes'  
order by usubjid;  
run;
```

```
data TSDTM.EX (drop=counter);  
    set work._EX;  
    by USUBJID;  
  
    if first.USUBJID then counter=0;  
    counter+1;  
    EXSEQ=COUNTER;  
run;
```

Code 4: Simplified SAS code to generate Exposure (EX) dataset.

STUDYID	DOMAIN	USUBJID	EXSEQ	EXREFID	EXTRT	EXDOSE	EXDOSU	EXDOSFRM	EXDOSFRQ	EXROUTE	EXSTDTC	EXENDTC	EXSTDY	EXENDY	EPOCH
BLCL-XXXX	EX	BLCL-XXXX_01	1	R	Reference Product	200 mg		TABLET, FILM COATED	ONCE	ORAL	2022-11-19T06:30	2022-11-19T06:30	1	1	TREATMENT
BLCL-XXXX	EX	BLCL-XXXX_02	1	T	Treatment Product	100 mg		TABLET	ONCE	ORAL	2022-11-19T06:32	2022-11-19T06:32	1	1	TREATMENT
BLCL-XXXX	EX	BLCL-XXXX_03	1	T	Treatment Product	100 mg		TABLET	ONCE	ORAL	2022-11-19T06:34	2022-11-19T06:34	1	1	TREATMENT
BLCL-XXXX	EX	BLCL-XXXX_04	1	R	Reference Product	200 mg		TABLET, FILM COATED	ONCE	ORAL	2022-11-19T06:36	2022-11-19T06:36	1	1	TREATMENT
BLCL-XXXX	EX	BLCL-XXXX_05	1	T	Treatment Product	100 mg		TABLET	ONCE	ORAL	2022-11-19T06:38	2022-11-19T06:38	1	1	TREATMENT
BLCL-XXXX	EX	BLCL-XXXX_06	1	R	Reference Product	200 mg		TABLET, FILM COATED	ONCE	ORAL	2022-11-19T06:40	2022-11-19T06:40	1	1	TREATMENT
BLCL-XXXX	EX	BLCL-XXXX_07	1	T	Treatment Product	100 mg		TABLET	ONCE	ORAL	2022-11-19T06:42	2022-11-19T06:42	1	1	TREATMENT
BLCL-XXXX	EX	BLCL-XXXX_08	1	R	Reference Product	200 mg		TABLET, FILM COATED	ONCE	ORAL	2022-11-19T06:44	2022-11-19T06:44	1	1	TREATMENT
BLCL-XXXX	EX	BLCL-XXXX_09	1	T	Treatment Product	100 mg		TABLET	ONCE	ORAL	2022-11-19T06:46	2022-11-19T06:46	1	1	TREATMENT
BLCL-XXXX	EX	BLCL-XXXX_10	1	T	Treatment Product	100 mg		TABLET	ONCE	ORAL	2022-11-19T06:48	2022-11-19T06:48	1	1	TREATMENT
BLCL-XXXX	EX	BLCL-XXXX_11	1	R	Reference Product	200 mg		TABLET, FILM COATED	ONCE	ORAL	2022-11-19T06:50	2022-11-19T06:50	1	1	TREATMENT
BLCL-XXXX	EX	BLCL-XXXX_12	1	T	Treatment Product	100 mg		TABLET	ONCE	ORAL	2022-11-19T06:52	2022-11-19T06:52	1	1	TREATMENT

Figure 7: Example of an output of SDTM Exposure (EX) domain dataset.

Adverse Events (AE)

An important event domain is Adverse Events (AE) which contains a description of medical occurrences in a subject throughout his involvement in a study. As seen in **Figure 8**, these events may or may not be related to treatment exposure. The collection of adverse events is sponsor defined and may vary. According to protocol definition, adverse events may be submitted as one record that encapsulates an event from start to end, or at a more granular level, in which a new record may appear as severity, causality or seriousness changes. In this case, AEs are reported as one record and the terms (AETERM) used are from a pre-specified list of terms. MedDRA (Medical Dictionary for regulatory Activities) is a highly specific hierarchical standardised terminology developed by the ICH. If the MedDRA dictionary was used to code events, then the levels of the hierarchy are presented in text and code form. For instance –LLT is used to represent the Lowest Level Term and –HLT for the Highest Level Term. The system organ class associated with the AE is represented by the –BODSYS variable. In this domain, the EPOCH in which the event occurs is derived from the start date and time of the AE.

USUBJID	ASEQ	AETERM	AELLT	AELLTCD	AEDECOD	AEPTCD	AEHLT
BLCL-XXXX_04	1	Left parietal headache	Parietal headache	10048969	Headache	10019211	Headaches NEC
BLCL-XXXX_04	2	Localized erythema left arm	Localized erythema	10064351	Erythema	10015150	Erythemas
BLCL-XXXX_07	1	Dry cough	Dry cough	10013773	Cough	10011224	Coughing and associated symptoms
BLCL-XXXX_07	2	Occipital headache	Occipital headache	10048968	Headache	10019211	Headaches NEC
USUBJID	AEHLGT	AEHLGTC	AEBODSYS	AEBDSYCD	AESOC	AESOC	AESOC
BLCL-XXXX_04	Headaches	10019231	Nervous system disorders	10029205	Nervous system disorders		10029205
BLCL-XXXX_04	Epidermal and dermal conditions	10014982	Skin and subcutaneous tissue disorders	10040785	Skin and subcutaneous tissue disorders		10040785
BLCL-XXXX_07	Respiratory disorders NEC	10038716	Respiratory, thoracic and mediastinal disorders	10038738	Respiratory, thoracic and mediastinal disorders		10038738
BLCL-XXXX_07	Headaches	10019231	Nervous system disorders	10029205	Nervous system disorders		10029205
USUBJID	AESER	AEACN	AEREL	EPOCH	AESTDTC	AEENDTC	
BLCL-XXXX_04	N	NOT APPLICABLE	NOT REASONABLY POSSIBLE	SCREENING	2022-11-16T18:00	2022-11-17T02:00	
BLCL-XXXX_04	N	NOT APPLICABLE	NOT REASONABLY POSSIBLE	TREATMENT	2022-11-19T09:30	2022-11-20T09:00	
BLCL-XXXX_07	N	NOT APPLICABLE	NOT REASONABLY POSSIBLE	SCREENING	2022-11-08T14:00	2022-12-03T20:00	
BLCL-XXXX_07	N	DOSE NOT CHANGED	REASONABLY POSSIBLE	TREATMENT	2022-11-19T16:45	2022-11-19T19:40	

Figure 8: Example of an output of SDTM Adverse Event (AE) domain dataset.

Pharmacokinetic Concentrations and Parameters (PC and PP)

The pharmacokinetics concentrations (PC) domain contains the concentrations of analytes and/or metabolites according to protocol determined timepoints. All required and expected variables constitute the dataset in **Figure 9** which gives an example of what the PC domain would look like for one subject. PCTESTCD and PCTEST, when uncensored would refer to the analyte or specimen name in question. To create this dataset it is necessary to receive the results from the biolaboratory assigned to perform the analysis on the blood samples collected. The biolaboratory will, usually, be blinded, even in simple open-label studies, and be external to the company, to avoid any biases. Variable PCORRES contains the exact result values that are received, while PCSTRESC and PCSTRESN are comprised of results in character and numeric format, accordingly. The following record and variable qualifiers are presented according to prior established Bioanalytical Plan and reported in Bioanalytical Report from the laboratory. These constitute, the specification of the specimen material (i.e. “Plasma”), the method that was employed to examine the samples, and lower and upper limit of quantification that determines when a result value is presented as “BLQ” or “ULQ”, respectively. It is important to have reference time point variables (PCRFDTC and PCTPTREF) for analysis interpretation and to relate records to their consequent analysis. These timepoints are in reference to dose administration.

STUDYID	DOMAIN	USUBJID	PCSEQ	PCTESTCD	PCTEST	PCCAT	PCORRES	PCORRESU	PCSTRESC	PCSTRESN	PCSTRESU	PCNAM
BLCL-XXXX	PC	BLCL-XXXX_02	1	XXXX	XXXX	ANALYTE	BLQ	ng/mL	BLQ		ng/mL	XXX INC.
BLCL-XXXX	PC	BLCL-XXXX_02	2	XXXX	XXXX	ANALYTE	112.9	ng/mL	112.9	112.90	ng/mL	XXX INC.
BLCL-XXXX	PC	BLCL-XXXX_02	3	XXXX	XXXX	ANALYTE	332.5	ng/mL	332.5	332.50	ng/mL	XXX INC.
BLCL-XXXX	PC	BLCL-XXXX_02	4	XXXX	XXXX	ANALYTE	421.0	ng/mL	421.0	421.00	ng/mL	XXX INC.
BLCL-XXXX	PC	BLCL-XXXX_02	5	XXXX	XXXX	ANALYTE	516.3	ng/mL	516.3	516.30	ng/mL	XXX INC.
BLCL-XXXX	PC	BLCL-XXXX_02	6	XXXX	XXXX	ANALYTE	1908.1	ng/mL	1908.1	1908.10	ng/mL	XXX INC.
BLCL-XXXX	PC	BLCL-XXXX_02	7	XXXX	XXXX	ANALYTE	906.8	ng/mL	906.8	906.80	ng/mL	XXX INC.
BLCL-XXXX	PC	BLCL-XXXX_02	8	XXXX	XXXX	ANALYTE	742.7	ng/mL	742.7	742.70	ng/mL	XXX INC.
BLCL-XXXX	PC	BLCL-XXXX_02	9	XXXX	XXXX	ANALYTE	333.6	ng/mL	333.6	333.60	ng/mL	XXX INC.
BLCL-XXXX	PC	BLCL-XXXX_02	10	XXXX	XXXX	ANALYTE	145.2	ng/mL	145.2	145.20	ng/mL	XXX INC.

USUBJID	PCSPEC	PCMETHOD	PCLLOQ	PCULLOQ	EPOCH	VISITNUM	VISIT	VISITDY	PCDTC	PCTPT	PCTPTNUM	PCRFDTC	PCTPTREF
BLCL-XXXX_02	PLASMA	HPLC/MS/MS	50	1000	TREATMENT	3.0	Confinement	1	2022-11-19T06:17	PRE-DOSE		0.0	Treatment Dose 2022-11-19T06:32
BLCL-XXXX_02	PLASMA	HPLC/MS/MS	50	1000	TREATMENT	3.0	Confinement	1	2022-11-19T07:32	1H POST-DOSE		1.0	Treatment Dose 2022-11-19T06:32
BLCL-XXXX_02	PLASMA	HPLC/MS/MS	50	1000	TREATMENT	3.0	Confinement	1	2022-11-19T08:32	2H POST-DOSE		2.0	Treatment Dose 2022-11-19T06:32
BLCL-XXXX_02	PLASMA	HPLC/MS/MS	50	1000	TREATMENT	3.0	Confinement	1	2022-11-19T08:52	2.3333H POST-DOSE		2.3	Treatment Dose 2022-11-19T06:32
BLCL-XXXX_02	PLASMA	HPLC/MS/MS	50	1000	TREATMENT	3.0	Confinement	1	2022-11-19T09:12	2.6667H POST-DOSE		2.7	Treatment Dose 2022-11-19T06:32
BLCL-XXXX_02	PLASMA	HPLC/MS/MS	50	1000	TREATMENT	3.0	Confinement	1	2022-11-19T09:32	3H POST-DOSE		3.0	Treatment Dose 2022-11-19T06:32
BLCL-XXXX_02	PLASMA	HPLC/MS/MS	50	1000	TREATMENT	3.0	Confinement	1	2022-11-20T06:32	24H POST-DOSE		24.0	Treatment Dose 2022-11-19T06:32
BLCL-XXXX_02	PLASMA	HPLC/MS/MS	50	1000	TREATMENT	4.0	Ambulatory	2	2022-11-20T18:32	36H POST-DOSE		36.0	Treatment Dose 2022-11-19T06:32
BLCL-XXXX_02	PLASMA	HPLC/MS/MS	50	1000	TREATMENT	4.0	Ambulatory	2	2022-11-21T06:32	48H POST-DOSE		48.0	Treatment Dose 2022-11-19T06:32
BLCL-XXXX_03	PLASMA	HPLC/MS/MS	50	1000	TREATMENT	4.0	Ambulatory	2	2022-11-22T06:34	72H POST-DOSE		72.0	Treatment Dose 2022-11-19T06:32

Figure 9: Example of an output of SDTM Pharmacokinetic Concentrations (PC) domain dataset.

After generating SDTM PC, the dataset is exported in “.xpt” SAS format so that the PK department can analyse the results according to parameters defined in Statistical Analysis Plan (SAP), for each record. Once PK analysis data is ready, it can be imported into the SAS project. The pharmacokinetics parameters (PP) domain is derived from PC data that has been analysed according to specific parameters. In this example, five parameters were considered (Figure 10).

STUDYID	DOMAIN	USUBJID	PPSEQ	PTESTCD	PTEST	PPORRES	PPORRESU	PPSTRESC	PPSTRESN	PPSTRESU	PPSPEC	PPRFTDTC
BLCL-XXXX	PP	BLCL-XXXX_02	1	AUC072	AUC from 0h to 72h	145029.3	h*ng/mL	145029.3	145029.3	h*ng/mL	PLASMA	2022-11-19T06:32
BLCL-XXXX	PP	BLCL-XXXX_02	2	CMAX	Max Conc	8664.2	ng/mL	8664.2	8664.2	ng/mL	PLASMA	2022-11-19T06:32
BLCL-XXXX	PP	BLCL-XXXX_02	3	LAMZ	Lambda z	0.023	/h	0.023	0.023	/h	PLASMA	2022-11-19T06:32
BLCL-XXXX	PP	BLCL-XXXX_02	4	LAMZHL	Half-Life Lambda z	26.6	h	26.6	26.6	h	PLASMA	2022-11-19T06:32
BLCL-XXXX	PP	BLCL-XXXX_02	5	TMAX	Time of CMAX	4.2	h	4.2	4.2	h	PLASMA	2022-11-19T06:32

Figure 10: Example of an output of SDTM Pharmacokinetic Parameters (PP) domain dataset.

2.4. ADaM Datasets

After ensuring that the SDTM datasets created contain all collected study data, ADaM datasets can be programmed. As mentioned in section 2.2.2, ADaM compiles ready for analysis data that can be traced back to SDTM. Before creating ADaM datasets it is important to have the specific analysis needs in mind and select the best suited ADaM data structure for that goal. The analysis datasets that I created are detailed below and include Subject-Level Analysis Dataset (ADSL), Pharmacokinetic Concentrations and Pharmacokinetics Parameters Analysis Datasets (ADPC and ADPP) and Adverse Events Analysis Datasets (ADAE). Ideally, the final dataset will be a DATA/SQL step and a PROC REPORT away and should allow a programmer to produce outputs with least amount of manipulation.

Subject-Level Analysis Dataset (ADSL)

The first dataset that is required for submission is the ADSL dataset. In simple terms, it is useful to create an ADSL dataset that will contain all data pertaining to demographic, exposure, disposition, and subject-level flags, that is relevant to subsequent datasets creations. Below is a code for the creation of FLAG variables (**Code 5**) and the merging with the predecessor SDTM DM dataset for a simplified ADSL output with only required variables (**Figure 11**). The resulting output highlights that not all subjects were randomized or completed the study, and therefore, might not be part of the safety analysis population.

```
proc sql;
  create table work._FLAGS as
    select distinct dm.usubjid,
      case when ex.usubjid~=" or ex2.usubjid~=" then 'Y' else 'N' end as SAFFL,
      case when ds.usubjid=" then 'Y' else 'N' end as COMPLFL,
      case when sdm.usubjid~=" then 'Y' else 'N' end as RANDFL
    from SDTM.DM as dm
  left join SDTM.EX as ex on (dm.usubjid=ex.usubjid and ex.exrefid='T')
  left join SDTM.EX as ex2 on (dm.usubjid=ex2.usubjid and ex2.exrefid='R')
  left join SDTM.DS as ds on (ds.usubjid=dm.usubjid and ds.dsdecod not in ('COMPLETED',
'INFORMED CONSENT OBTAINED'))
  left join SDTM.SUPPDM as sdm on (sdm.usubjid=dm.usubjid and sdm.qnam='RANDNO');
run;

proc sql;
  create table adam.ADSL as
  select  dm.studyid,
          dm.usubjid,
          dm.subjid as SUBJID,
          dm.siteid,
          dm.age,
          dm.ageu,
          dm.sex,
          dm.race,
          fl.SAFFL,
          fl.COMPLFL,
          fl.RANDFL,
          dm.arm,
          dm.actarm,
          case when dm.armcd ~=" then dm.arm else " end as TRT01P,
          case when dm.actarmcd ~=" then dm.actarm else " end as TRT01A,
          input(ex.exstdtc, e8601dt.) format=E8601DT. as TRTSDTM,
          input(ex.exendtc, e8601dt.) format=E8601DT. as TRTEDTM,
          'S' as TRTSTMF,
          'S' as TRTETMF,
```

```

                                datepart(input(sv.svstddtc, e8601dt.)) format=yymmdd10. as RANDDT
from SDTM.DM as dm
left join work._FLAGS as fl on (fl.usubjid=dm.usubjid)
left join SDTM..EX as ex on (ex.usubjid=dm.usubjid)
left join SDTM.SV as sv on (sv.usubjid=dm.usubjid and sv.visit='Confinement');
run;

```

Code 5: Simplified SAS code for generating ADaM Subject-Level (ADSL) dataset.

STUDYID	USUBJID	SUBJID	SITEID	AGE	AGEU	SEX	RACE	SAFFL	COMPLFL	RANDFL
BLCL-XXXX	BLCL-XXXX_01	0012	BlueClinical Phase I	28	YEARS	F	WHITE	Y	Y	Y
BLCL-XXXX	BLCL-XXXX_02	0084	BlueClinical Phase I	38	YEARS	F	BLACK OR AFRICAN AMERICAN	Y	Y	Y
BLCL-XXXX	BLCL-XXXX_03	0015	BlueClinical Phase I	25	YEARS	M	WHITE	Y	Y	Y
BLCL-XXXX	BLCL-XXXX_04	0065	BlueClinical Phase I	27	YEARS	F	WHITE	Y	Y	Y
BLCL-XXXX	BLCL-XXXX_05	0086	BlueClinical Phase I	36	YEARS	M	WHITE	Y	Y	Y
BLCL-XXXX	BLCL-XXXX_06	0022	BlueClinical Phase I	45	YEARS	M	WHITE	Y	Y	Y
BLCL-XXXX	BLCL-XXXX_07	0040	BlueClinical Phase I	30	YEARS	M	WHITE	Y	Y	Y
BLCL-XXXX	BLCL-XXXX_08	0048	BlueClinical Phase I	21	YEARS	M	WHITE	Y	Y	Y
BLCL-XXXX	BLCL-XXXX_09	0021	BlueClinical Phase I	22	YEARS	F	BLACK OR AFRICAN AMERICAN	Y	Y	Y
BLCL-XXXX	BLCL-XXXX_10	0089	BlueClinical Phase I	31	YEARS	F	WHITE	Y	Y	Y
BLCL-XXXX	BLCL-XXXX_11	0090	BlueClinical Phase I	31	YEARS	M	WHITE	Y	Y	Y
BLCL-XXXX	BLCL-XXXX_12	0006	BlueClinical Phase I	28	YEARS	F	WHITE	Y	Y	Y
BLCL-XXXX	SCR_0035	0035	BlueClinical Phase I	29	YEARS	M	BLACK OR AFRICAN AMERICAN	N	N	N

Figure 11: Example of an output of ADaM Subject-Level (ADSL) dataset.

Pharmacokinetic Concentrations and Pharmacokinetics Parameters Analysis Datasets (ADPC and ADPP)

ADPC and ADPP follow a basic data structure with SDTM findings PC and PP domains as predecessor. The implementation of these datasets is based on rules defined in SAP to create pharmacokinetic graphs that translate the curve of concentrations on a time axis.

One common rule relates to how certain concentration values should be transformed when they are below the lower limit of quantification (LLOQ) depending on the timepoint of blood sampling. In the case of the following SAP specification: “For plotting data in a semi-logarithmic scale, before the time (T_{max}) it takes for the drug to reach the maximum concentration (C_{max}), concentrations below the LLOQ will be considered $\frac{1}{2}$ of LLOQ and after T_{max} , concentrations below the LLOQ will be considered as missing”, the steps in **Code 6** would translate that rule.

```

/* create column with tmax values */
data work._rule1(keep=usubjid tmax);
  set tsdtm.pp(where=(PPTTESTCD='TMAX'));
  tmax=input(pporres,d.);
run;

/*merge with concentrations (PC) dataset*/
proc sql;
  create table work._rule1a as
  select pc.*,
         r.tmax,
         case when pc.pcorres ~='BLQ' then input(pc.pcorres, best.) else 0 end as pcorresn
  from tsdtm.pc as pc
  inner join work._rule1 as r on (r.usubjid=pc.usubjid);
run;

data work._rule1b;
  set work._rule1a;

  /* if tmax wasn't reached, transform concentrations to 1/2 of LLOQ */
  if tmax>pctptnum and pcorresn<&_LLOQ then pcorres=strip(put(&_LLOQ/2,8.2));
  /* if timepoint is after tmax, then transform results to 0*/
  if tmax<pctptnum and pcorresn<&LLOQ then pcorres= "Missing";
run;

```

Code 6: Simplified SAS code for specific SAP specified rules.

The transformed result variables would constitute the AVAL and AVALC columns of a typical ADPC dataset and DTYPE variable will contain information on the type of derivation for that record (**Figure 12**). The DTYPE variable must be populated according to a specific controlled terminology codelist with values such as “BC” (best case scenario), “LOCF” (last observation carried forward), amongst others. When no imputation or modification has been made, the variable will be null.

STUDID	USUBJID	SUBJID	SITEID	TRTA	TRTP	AVISIT	AVISITN	ATPT	ATPTN	PARAM	PARAMCD	AVAL	AVALC	DTYPE	SRCDOM	SRCVAR	SRCSEQ
BLCL-XXX	BLCL-XXXX_02	0084	BlueClinical Phase I	Test Product	Test Product	Confinement	3.0	PRE-DOSE		0.0 XXXX (ng/mL)	XXXX	0.00	0	BC	PC	PCSTRESN	1
BLCL-XXX	BLCL-XXXX_02	0084	BlueClinical Phase I	Test Product	Test Product	Confinement	3.0	1H POST-DOSE		1.0 XXXX (ng/mL)	XXXX	112.90	112.9		PC	PCSTRESN	2
BLCL-XXX	BLCL-XXXX_02	0084	BlueClinical Phase I	Test Product	Test Product	Confinement	3.0	2H POST-DOSE		2.0 XXXX (ng/mL)	XXXX	332.50	332.5		PC	PCSTRESN	3
BLCL-XXX	BLCL-XXXX_02	0084	BlueClinical Phase I	Test Product	Test Product	Confinement	3.0	2.3333H POST-DOSE		2.3 XXXX (ng/mL)	XXXX	421.00	421.0		PC	PCSTRESN	4
BLCL-XXX	BLCL-XXXX_02	0084	BlueClinical Phase I	Test Product	Test Product	Confinement	3.0	2.6667H POST-DOSE		2.7 XXXX (ng/mL)	XXXX	516.30	516.3		PC	PCSTRESN	5
BLCL-XXX	BLCL-XXXX_02	0084	BlueClinical Phase I	Test Product	Test Product	Confinement	3.0	3H POST-DOSE		3.0 XXXX (ng/mL)	XXXX	1908.10	1908.1		PC	PCSTRESN	6
BLCL-XXX	BLCL-XXXX_02	0084	BlueClinical Phase I	Test Product	Test Product	Confinement	3.0	24H POST-DOSE		24.0 XXXX (ng/mL)	XXXX	906.80	906.8		PC	PCSTRESN	7
BLCL-XXX	BLCL-XXXX_02	0084	BlueClinical Phase I	Test Product	Test Product	Ambulatory	4.0	36H POST-DOSE		36.0 XXXX (ng/mL)	XXXX	742.70	742.7		PC	PCSTRESN	8
BLCL-XXX	BLCL-XXXX_02	0084	BlueClinical Phase I	Test Product	Test Product	Ambulatory	4.0	48H POST-DOSE		48.0 XXXX (ng/mL)	XXXX	333.60	333.6		PC	PCSTRESN	9
BLCL-XXX	BLCL-XXXX_02	0084	BlueClinical Phase I	Test Product	Test Product	Ambulatory	4.0	72H POST-DOSE		72.0 XXXX (ng/mL)	XXXX	145.20	145.2		PC	PCSTRESN	10

Figure 12: Example of an output of ADaM Plasma Concentration (ADPC) dataset.

As referenced, SAP might also specify certain parameter derivations needed to output different pharmacokinetics curves. In the case of a semi-logarithmic scale, the log values can constitute a new column in the ADPP dataset (**Figure 13**).

STUDYID	USUBJID	SUBJID	SITEID	TRTP	TRTA	PARAM	PARAMC	PARCAT1	AVAL	AVALC	LOG	SRCDOM	SRCVAR	SRCSEQ
BLCL-XXXX	BLCL-XXXX_02	0084	BlueClinical Phase I	Test Product	Test Product	Time of CMAX (h)	TMAX	XXXX	4.200	4.20		PP	PPSTRESN	1
BLCL-XXXX	BLCL-XXXX_02	0084	BlueClinical Phase I	Test Product	Test Product	Half-Life Lambda z (h)	LAMZHL	XXXX	26.600	26.6		PP	PPSTRESN	2
BLCL-XXXX	BLCL-XXXX_02	0084	BlueClinical Phase I	Test Product	Test Product	Lambda z (/h)	LAMZ	XXXX	0.023	0.036		PP	PPSTRESN	3
BLCL-XXXX	BLCL-XXXX_02	0084	BlueClinical Phase I	Test Product	Test Product	AUC from 0h to 72h (h*ng/mL)	AUC072	XXXX	145029.300	145029.3	5.16145	PP	PPSTRESN	4
BLCL-XXXX	BLCL-XXXX_02	0084	BlueClinical Phase I	Test Product	Test Product	Max Conc (ng/mL)	CMAX	XXXX	8664.200	8664.2	3.93772	PP	PPSTRESN	5

Figure 13: Example of an output of ADaM Pharmacokinetic Parameters (ADPP) dataset.

Adverse Events Analysis Datasets (ADAE)

ADAE constitutes an OCCDS structure and compiles all information necessary for practical AE reporting. It will trace back to all SDTM AE variables and join additional information that is stored in SUPPAE. The resulting dataset (**Figure 14**) typically has relevant AE flags, such as those that indicate if AE was treatment emergent (TRTEMFL) or if it occurred before treatment administration (PREFL).

STUDYID	USUBJID	SUBJID	AESQ	AETERM	TRTEMFL	PREFL	SAFFL	AESTDTC	ASTDT	ASTDTM	ASTTMF
BLCL-XXXX	BLCL-XXXX_04	0065	1	Left parietal headache		Y	Y	2022-11-16T18:00	2022-11-16	2022-11-16T18:00:00	S
BLCL-XXXX	BLCL-XXXX_04	0065	2	Localized erythema left arm	Y		Y	2022-11-19T09:30	2022-11-19	2022-11-19T09:30:00	S
BLCL-XXXX	BLCL-XXXX_07	0040	2	Dry cough		Y	Y	2022-11-08T14:00	2022-11-08	2022-11-08T14:00:00	S
BLCL-XXXX	BLCL-XXXX_07	0040	1	Occipital headache	Y		Y	2022-11-19T16:45	2022-11-19	2022-11-19T16:45:00	S

Figure 14: Example of an output of ADaM Adverse Event (ADAE) dataset.

2.5. Tables, Figures and Listings (TFLs)

To report the data that has been collected, organized, and analysed previously, in a readable concise way, clinical data programmers generate TFLs that will be an essential component of the Clinical Study Report (CSR). All TFLs that must be generated are defined in the Statistical Analysis Plan (SAP), that is thoroughly reviewed with input from different departments, such as CDP, PK, CDM, MW and QA (Quality Assurance).

Tables serve the purpose of summarizing key aspects of the study by means of descriptive statistics (i.e. mean, median, maximum, minimum and standard deviation). Listings are more detailed and provide information that might be organized by treatment group or a specific population flag. Figures are a visual representation of data, such as graphs that represent the pharmacokinetics of a drug.

During my internship I generated numerous, SAP specified TFLs for a study. Apart from data manipulation steps to summarise or list the data, there is an important SAS procedure employed, the PROC REPORT step. This procedure specifies what columns will constitute the table or listing, and defines how they will be displayed and formatted, including order, font size, alignment, width, and other customizable options. The output is shown in **Figure 15**, where Treatment Emergent Adverse Events are listed.

Subject No.	Investigational Product	Reported Term	Serious?	Causality	Action Taken	Adverse Event Start Date and Time	Adverse Event End Date and Time
BLCL-XXXX_04	Reference Product	Lumbar pain	NON-SERIOUS	NOT REASONABLY POSSIBLE	NOT APPLICABLE	19NOV2022 09:30	20NOV2022 09:00
BLCL-XXXX_07	Test Product	Occipital headache	NON-SERIOUS	REASONABLY POSSIBLE	DOSE NOT CHANGED	19NOV2022 16:45	19NOV2022 19:40

Figure 15: Example of Treatment-Emergent Listing for a Clinical Study Report.

2.6. Pinnacle validation

Pinnacle 21 Validator is the leading industry tool for validating SDTM and ADaM datasets against CDISC and FDA requirements³³. Through uploading the outputted SDTM and ADaM datasets into the validator, it returns a report on dataset content and metadata, in excel format. The Excel includes four worksheets:

- General Summary: a listing of validated datasets and the errors, warnings, and notices for each.
- Issue Summary: a listing of issues grouped by dataset, categorized by severity.
- Details: information for each reported issue instance, including the dataset name, number of records affected with the issue, and values of the affected variables.
- Rules: a list of validation rules.

This is a great way to validate the programming outputs and comb through possible errors and issues before advancing with data reporting.

2.7. Internal Quality Control

Throughout the pipeline of CDP activities there are several quality control timepoints that are designated to ensure work quality and reliability. The data that is being handled is, not only sensitive, but vast, therefore, it's good practice for another member of the team to perform QC on the tasks that were done.

This practice encouraged me to look closely at others work and brought a better understanding of the tasks at hand. Moreover, it led to the development of critical problem solving in ways to improve efficiency in data validation without sacrificing accuracy. This was done by applying knowledge in other programming languages, namely R, to optimize this process. A useful discovery was “comparedf” function from the R package “arsenal” that is designed to compare two data frames and identify any discrepancies, in detail. The output between the comparison of the datasets for submission and the datasets generated in R must be stored in a proper log, for transparency and audit trail. This can be achieved through package “logr” that stores exact logging date and time, functions, and results.

IV. CONCLUDING REMARKS

Assuming the role of a clinical data programmer has been a valuable and enriching experience that has allowed me to develop my technical skills in the programming world and build upon clinical knowledge gained in my master's degree.

The range of activities that CDP is a part of, has given me a deeper insight into the intricacies of clinical studies and the regulatory requirements for their submission. The emphasis of this report is CDISC compliance and depicts the level of rigor, accuracy, and transparency of clinical data handling. It also highlights the ever learning and critical thinking nature of what it means to program in a highly regulated and standardized field. There is a need for both technical expertise and the willingness to learn and know conformance rules.

Undertaking this internship set a new course for my professional life and gave me the opportunity to face weaknesses and recognise strengths. Although I am a fast learner and can easily adapt to new environments, there have been personal challenges that came from being in a professional setting. The everyday responsibilities taught me new organizational skills and the importance of communicating my doubts and mistakes. Working at BlueClinical also allowed me to further develop team skills by understanding that every task is dependent on the synergy and communication between team members and other departments. When it comes to hard skills, I learned that only with time and experience can someone become proficient in their role. Throughout the nine months of the internship, I was able to accomplish the objectives that were set for me and be a part of the CDP pipeline, namely being able to randomize trials, generate SDTM and ADaM datasets for regulatory submission, and create TLFs for CSR integration. The CDP role goes further and there is much more to learn, and I am ready for new challenges.

Overall, this internship provided me with the experience of exploring the intersection of two fields I have great interest in – programming and clinical work. Through this opportunity I was able to learn, adapt to new challenges, and expand my understanding of these two areas.

Bibliography

1. BlueClinical. No Title. Accessed April 6, 2023. <https://www.blueclinical.com/>
2. BlueClinical - About Us. Accessed April 6, 2023. <https://www.blueclinical.com/about-us>
3. BlueClinical. Translational Medicine Consultancy. Accessed April 6, 2023. <https://www.blueclinical.com/services/pre-clinical-development>
4. Committee for Human Medicinal Products Guideline for Good Clinical Practice E6(R2) Step 5.; 2016. Accessed April 8, 2023 https://www.ema.europa.eu/en/documents/scientific-guideline/ich-guideline-good-clinical-practice-e6r2-step-5_en.pdf.
5. BlueClinical. Therapeutic Endpoint Studies. Accessed April 6, 2023. <https://www.blueclinical.com/services/clinical-research-partnership>
6. BlueClinical. Bioequivalence and Phase I Studies. Accessed April 6, 2023. <https://www.blueclinical.com/services/fase-i>
7. Junod SW. FDA and Clinical Drug Trials : A Short History. *A Quick Guid to Clin Trials*. Published online 2008:22-55.
8. Emanuel EJ, Wendler D, Grady C. What Makes Clinical Research Ethical ? 2000;283(20):2701-2711.
9. Mandal J, Acharya S, Parija SC. Ethics in human research. *Trop Parasitol*. 2011;1(1):2-3. doi:10.4103/2229-5070.72105.
10. Brandt AM. Racism and Research : The Case of the Tuskegee Syphilis Study The of Tuskegee Syphilis Study. 1978;8(6):21-29.
11. Yousuf Guraya S, London NJM, Salman Guraya S. Ethics in medical research. *J Microsc Ultrastruct*. 2014;2(3):121-126. doi:10.1016/j.jmau.2014.03.003
12. 1. WMA - The World Medical Association-WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects. Wma.net. Published 2014. Accessed May 11, 2023. <https://www.wma.net/policies-post/wma-declaration->

of-helsinki-ethical-pr.

13. Vijayanathan A, Nawawi O. The importance of Good Clinical Practice guidelines and its role in clinical trials. Published online 2008. doi:10.2349/bijj.4.1.e5
14. INFARMED. About Infarmed. Accessed April 20, 2023. <https://www.infarmed.pt/web/infarmed-en/about-infarmed>
15. CEIC. Missão e Visão. Accessed April 20, 2023. <https://www.ceic.pt/missao>
16. CNPD. Histórico da CNPD. Accessed April 20, 2023. <https://www.cnpd.pt/cnpd/o-que-somos-e-quem-somos/historico-da-cnpd/>
17. EMA. Who we are. Accessed April 20, 2023. <https://www.ema.europa.eu/en/about-us/who-we-are>
18. FDA. About FDA. Accessed April 20, 2023. [fda.gov/about-fda](https://www.fda.gov/about-fda)
19. FDA. Study Data Standards Resources. Accessed March 21, 2023. <https://www.fda.gov/industry/fda-data-standards-advisory-board/study-data-standards-resources>
20. Bhatt A. Evolution of clinical research: a history before and beyond james lind. *Perspect Clin Res.* 2010;1(1):6-10. 2010;1(1).
21. Souza T, Kush R, Evans JP. Global clinical data interchange standards are here ! *2007;12(February)*. doi:10.1016/j.drudis.2006.12.012
22. CDISC. CDISC. Accessed April 26, 2023. <https://www.cdisc.org/>
23. Wood F, Guintier T. Evolution and Implementation of the CDISC Study Data Tabulation Model (SDTM). *2008;1(1):20-28*. doi:10.1179/175709208X334623
24. CDISC. CDASH. Accessed April 26, 2023. <https://www.cdisc.org/standards/foundational/cdash>
25. CDISC. PRM. Accessed April 26, 2023. <https://www.cdisc.org/standards/foundational/protocol>
26. Krbick WR, Carolina N. Toward a Comprehensive CDISC Submission Data Standard. Published online 2007:373-382.

27. CDISC. SEND. Accessed April 26, 2023.
<https://www.cdisc.org/standards/foundational/send>
28. CDISC. SDTM. Accessed May 4, 2023.
<https://www.cdisc.org/standards/foundational/sdtm>
29. CDISC. ADAM. Accessed May 4, 2023.
<https://www.cdisc.org/standards/foundational/adam>
30. CDISC. A Short History of CDISC and SAS Transport Files. Accessed May 4, 2023. <https://www.cdisc.org/kb/articles/short-history-cdisc-and-sas-transport-files>
31. Kang M, Ragan BG, Park JH. Issues in outcomes research: An overview of randomization techniques for clinical trials. *J Athl Train*. 2008;43(2):215-221. doi:10.4085/1062-6050-43.2.215
32. Suresh K. An overview of randomization techniques: An unbiased assessment of outcome in clinical research. *J Hum Reprod Sci*. 2011;4(1):8-11. doi:10.4103/0974-1208.82352
33. CERTARA. Pinnacle 21 Enterprise CDISC Software. Accessed May 5, 2023.
https://www.certara.com/pinnacle-21-enterprisesoftware/?utm_medium=ppc&utm_source=googleads&utm_content=pediatrics&utm_campaign=nonbrandcdisc&gclid=Cj0KCQjwpPKiBhDvARIsACn-gzCY-YUmnFARC4hu-0IbuUII8Nt9PNNFkoqa9fknAhQkn6y5zeM9wdQaAs7CEALw_wcB&gclsrc=aw.ds