



Universidade de Aveiro
Ano 2023

**João Tiago Pereira De
Oliveira**

**PROGRAMAÇÃO DE RELATÓRIOS DE DADOS
CLÍNICOS PARA SUBMISSÃO REGULAMENTAR DE
ENSAIOS CLÍNICOS DE PHASE I**

**PROGRAMMING OF CLINICAL DATA REPORTS FOR
REGULATORY SUBMISSION OF PHASE I CLINICAL
TRIALS**



**João Tiago Pereira De
Oliveira**

**PROGRAMAÇÃO DE RELATÓRIOS DE DADOS
CLÍNICOS PARA SUBMISSÃO REGULAMENTAR DE
ENSAIOS CLÍNICOS DE PHASE I**

**PROGRAMMING OF CLINICAL DATA REPORTS FOR
REGULATORY SUBMISSION OF PHASE I CLINICAL
TRIALS**

Relatório de estágio curricular apresentado à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Bioinformática Clínica, realizada sob a orientação científica da Professora Doutora Carla Alexandra Pina da Cruz Nunes, Professora Auxiliar do Departamento de Ciências Médicas da Universidade de Aveiro.

“Try and leave this world a little better than you found it, and when your turn comes to die, you can die happy in feeling that at any rate, you have not wasted your time but have done your best.”

Robert Baden-Powell

o júri

presidente

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

vogal – arguente principal

Prof. Doutor Pedro Miguel Ferreira de Sá Couto
Professor auxiliar da Universidade de Aveiro

vogal – orientadora

Prof^a. Doutora Carla Alexandra Pina da Cruz Nunes
Professora auxiliar da Universidade de Aveiro

agradecimentos

Ao Eng. Bernardo Almeida, pela oportunidade de estágio na BlueClinical.
À Verónica Mendonça, minha orientadora no local de estágio, por toda a formação que me transmitiu bem como pelo apoio dado durante todo o estágio.

À Prof^a Alexandra Nunes, minha orientadora da universidade, pelo seu apoio e disponibilidade na fase final do meu estágio.

À equipa de programação de dados clínicos, não só pelo caloroso acolhimento como pelo contributo no meu percurso profissional como programador de dados clínicos.

Aos meus colegas de licenciatura e de mestrado, pelo crescimento conjunto ao longo destes cinco anos.

A toda a minha família e amigos pelo apoio incondicional durante todo o meu percurso académico, destacando a minha Avó Alice, que infelizmente nos deixou neste último ano.

palavras-chave

Investigação Clínica, Estudos Clínicos, BlueClinical, Programação de Dados Clínicos, Standardização de Dados, TFLs.

resumo

O presente relatório descreve as atividades realizadas durante o estágio curricular em programação de dados clínicos desenvolvido na BlueClinical, Lda, com maior foco nas metodologias e questões regulamentares relacionadas com as Tabelas, Figuras e Listagens (TFLs) que irão ser representadas no Relatório do estudo Clínico (CSR), no âmbito do Mestrado em Bioinformática Clínica. O relatório começa com a descrição da empresa e das atividades específicas de um programador de dados clínicos. De seguida aborda aspetos fundamentais da Investigação Clínica, dando depois ênfase à sua situação em Portugal e ao ambiente regulamentar em que se encontra, como também aborda documentos associados a estudos clínicos a serem considerados nas TFLs. Este relatório também elabora de forma sucinta a standardização de dados, mencionando especificamente os modelos Study Data Tabulation Model (SDTM) e Analysis Data Model (ADaM), e alguns exemplos relevantes de cada modelo. Por último, as TFLs são explicadas, e de seguida exemplos de processos da criação de uma Tabela específica e de uma Figura específica serão abordados. Estas atividades referidas permitiram a aquisição de conhecimentos práticos em programação de dados clínicos.

keywords

Clinical Research, Clinical Studies, BlueClinical, Clinical Data Programming, Data Standardization, TFLs.

abstract

The following report describes the activities performed during the internship in clinical data programming developed at BlueClinical, Ltd, with a major focus on the methodologies and regulatory issues related to the Tables, Figures and Listings (TFLs) that will be represented in the Clinical Study Report (CSR), in the scope of the Master in Clinical Bioinformatics. The report begins with a description of the company and the specific activities of a clinical data developer. It then covers fundamental aspects of Clinical Research, with emphasis on its situation in Portugal and the regulatory environment it finds itself in, as well as documents associated with clinical studies to be considered in TFLs. This report also briefly elaborates on data standardization, specifically mentioning the Study Data Tabulation Model (SDTM) and Analysis Dataset Model (ADaM), and some relevant examples of each. Finally, TFLs are explained, and examples of the processes of creating a specific Table and a specific Figure are addressed. These referred activities allowed the acquisition of practical knowledge in clinical data programming.

Table of Contents

Abbreviation List	2
Figures List.....	4
Tables and Listing List.....	4
Code List	5
1. Introduction	6
1.1. BlueClinical Ltd.....	6
1.2. Clinical Data Programming (CDP)	7
1.3. Objectives	9
2. State of the Art.....	9
2.1. What is Clinical Research?	9
2.2. Regulatory landscape	14
2.3. Clinical Study Documents to be considered for TFLs.....	22
3. Data Standardization.....	23
3.1. STDM Datasets.....	23
3.2. ADaM Datasets	31
4. TFLs.....	34
4.1. Creation of a Table.....	36
4.2. Creation of a Figure	42
5. Discussion.....	44
6. Conclusion	48
7. Bibliography	48

Abbreviation List

ADaM	Analysis Data Model
ADR	Adverse Drug Reaction
ADAE	Adverse Events Analysis Dataset
ADPC	Pharmacokinetics Concentrations Analysis Dataset
ADSL	Analysis Data Subject Level
BDS	Basic Data Structure
CDISC	Clinical Data Interchange Standards Consortium
CDP	Clinical Data Programmer
CEIC	Comissão de Ética para a Investigação Clínica
CFR	Case Report Form
CRO	Clinical Research Organization
CRP	Clinical Research Partnership
CSP	Clinical Study Protocol
CSR	Clinical Study Report
CTA	Clinical Trial Applications
CTD	Common Technical Document
CTIS	Clinical Trials Information System
DM	Clinical Data Management
DMC	Data Monitoring Committees
eCTD	Electronic Common Technical Document
EDC	Electronic Data Capture
EEA	European Economic Area
EMA	European Medicines Agency
EWG	Expert Working Group
FDA	U.S. Food and Drug Administration
GCP	Good Clinical Practice
GDPR	General Data Protection Regulation

GVP	Good Pharmacovigilance Practice
ICH	International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use
MW	Medical Writing
OCCDS	Occurrence Data Structure
PK	Pharmacometric Department
SAP	Statistical Analysis Plan
SAS	Statistical Analysis System
SDTM	Study Data Tabulation Model
SiNATS	Sistema Nacional de Avaliação de Tecnologias de Saúde
TEAE	Treatment Emergent Adverse Event
TFL	Tables, Figures and Listings
WHO	World Health Organization

Figures List

Figure 1 - Clinical trials in Portugal between the 2 nd semester of 2005 and 2022. The dark blue bars represent the number of Clinical Trials submitted, the light blue bars represent the number of Clinical Trials authorized, and the gray bars represent the number of Clinical Trials not authorized. The orange line represents the average time in calendar days for authorization. Plot adapted from the 2005 to 4Q2022 Activity Indicators Statistics made available by INFARMED [11].	12
Figure 2 - Percentages of Clinical Trial applications submitted to INFARMED, divided by Clinical Development Phase. Phase I applications are indicated as green, Phase II applications are indicated as orange, Phase III applications are indicated as blue, and Phase IV applications are indicated as yellow. Plot adapted from the 2005 to 4Q2022 Activity Indicators Statistics made available by INFARMED [11].	12
Figure 3 - Simplified Example of a SV dataset.	26
Figure 4 - Simplified Example of a DM dataset.	27
Figure 5 - Simplified Example of a AE dataset.	29
Figure 6 - Simplified Example of a EX dataset.	30
Figure 7 - Simplified Example of a PC dataset.	31
Figure 8 - Simplified Example of an ADSL dataset.	32
Figure 9 - Simplified Example of an ADPC dataset.	33
Figure 10 - Simplified Example of an ADAE dataset.	34
Figure 11 - Arithmetic Mean Plasma Concentration Versus Time Profiles Following Administration of Analytes A and B.	44

Tables and Listing List

Listing 1 - Individual Pharmacokinetic Concentration Data	35
Listing 2 - Individual Adverse Events Concentration Data	35
Table 1 - Summary of TEAEs	42
Table 2 - Training performed during internship	45
Table 3 - Tasks performed during internship	47

Code List

Code 1 - Association of TEAEs with the study treatments.....	36
Code 2 - Number of Subjects in each treatment.....	37
Code 3 - Portion of query that summarizes the TEAEs regarding each study treatment.	37
Code 4 - Portion of query that summarizes the TEAEs in total.....	38
Code 5 - Portion of query that summarizes the TEAEs regarding their severity.....	38
Code 6 - Merging of each summarization dataset into one.....	39
Code 7 - Simplified query for the creation of the TFLSTYLE custom format style.	39
Code 8 - Portion of the query used for the report of the summarized TEAEs and exportation to a RTF file.....	41
Code 9 - Calculation of arithmetic mean plasma concentrations.....	43
Code 10 - Simplified Query used for Creation of the chart that presents the arithmetic mean plasma concentrations and exportation to RTF file.	43

1. Introduction

1.1. BlueClinical Ltd.

BlueClinical Ltd. was established on May 8, 2012, as a privately-owned company with the aim of offering scientific advice and engaging in research and development (R&D) activities for various healthcare products such as drugs, medical devices, and diagnostics. The company is particularly skilled in performing bioequivalence and early phase studies in both healthy and selected ill individuals, and also provides consulting services for translational medicine studies, and manages material and human resources. At the time of this report, BlueClinical operates through three business units, namely R&D, Phase I, and Clinical Research Partnership (CRP) [1].

The Phase I Clinical Pharmacology Unit, situated at Hospital da Prelada in Oporto, Portugal, specializes in conducting phase I clinical trials for healthy individuals or targeted populations. The unit manages the entire process from designing the study protocol to producing the final study report. With 61 beds, the unit has an added advantage of being located in a hospital, providing access to essential services such as emergency care. The Phase I unit performs various types of studies for both generic and innovative drugs in the global market. These studies include bioavailability/bioequivalence in healthy individuals or patients, as per the directives of regulatory authorities such as EMA, FDA (United States), or others. Additionally, the unit performs drug interaction, food effect, dose proportionality, first-in-human, and special population studies as part of Phase I clinical trials [1,2].

The R&D unit specializes in providing consulting services for drug development and full-service Clinical Research Organization (CRO) functions primarily for early phase trials involving patients. The R&D unit offers an extensive scope of services, which include consulting services that involve the preparation of scientific and regulatory opinions for regulatory authorities such as the FDA, EMA, and other national regulatory authorities. In addition, the R&D unit provides various services such as pharmacovigilance, clinical monitoring, medical management, project management, medical writing, pharmacometrics, regulatory affairs, and data management services, which are included in the complete clinical trial package or can be contracted separately [1,2].

The CRP unit is responsible for supporting clinical research activities in partner health institutions, with the goal of promoting efficiency, development, and excellence. Currently, there are twelve institutions that are part of the CRP network, which are the *Unidade Local de Saúde do Alto*

Minho (ULSAM), Hospital da Senhora da Oliveira Guimarães (HSOG), Unidade Local de Saúde de Matosinhos (ULSM), Centro Hospitalar de Vila Nova de Gaia/Espinho (CHVNG/E), Centro Hospitalar do Baixo Vouga (CHBV), Centro Hospitalar Leiria (CHL), Hospital Garcia de Orta (HGO), Centro Hospitalar Conde de Ferreira (CHCF), Unidade Local de Saúde da Guarda (ULSG), Caledial: Centro de Hemodiálise de Gaia, S.A., Hospital da Prelada (Partnership activation in 2022), and Hospital pediátrico da CHUC (Partnership Initiation in 2023). The CRP unit works collaboratively with these institutions to ensure their clinical research activities are conducted effectively and to the highest standards [1,2].

The curricular internship performed at BlueClinical started at 04/10/2022 and ended at 06/06/2023, corresponding to 1296 working hours, 650 of which occurred at the company.

1.2. Clinical Data Programming (CDP)

During the curricular internship, the activities performed were bound to the Clinical Data Programming and Biostatistics Department. These activities include but are not limited to: Randomization, External Data Integration, Data Standardization and Tables, Figures and Listings (TFLs).

Randomization is a process where participants are assigned to different treatment groups in an unbiased way. Its importance is due to the fact that it helps to ensure that the study results are valid and that any observed treatment effects are not due to biases or other confounding factors [3]. To avoid any kind of imbalance, regulatory agencies recommend that the randomization procedure be done block-wise. The process of block randomization involves the random assignment of participants to different treatments within specific blocks, ensuring an equal number in each group. For instance, if we consider a block size of 4, there are a total of 6 potential ways to evenly allocate participants into a block. The allocation process involves randomly selecting one of these possible orders and assigning the subsequent block of participants to the study groups based on the predetermined sequence. [3].

Here in BlueClinical, the randomization is planned according with information in the Clinical Study Protocol (CSP), and then code is prepared in order to generate a randomization list that will be distributed to the required departments. When it comes to blinded studies, randomization might require different steps, which depends on what was written in the CSP and was agreed with the Sponsor. In these cases, the step of Emergency Unblinding is usually described, most specifically its type and the situations where the unblinding can be done.

Overall process of External Data Integration consists in importation of data collected from different departments apart from CDP like the Data Management (DM) department and Pharmacometrics (PK) department or other entities that are not part of BlueClinical, such as the laboratories where bioanalysis is performed. External data typically includes data from biolaboratory tests, such as plasma concentrations and bioanalytical information, electronic case report forms (eCRFs) containing information on patients' data over the course of a study, pharmacokinetic and pharmacodynamic data that relates to the concentrations of drugs on the body and safety laboratory data, which includes reference values for hematology, clinical chemistry, and viral serology.

Data Standardization to SDTM and ADaM formats, which is one of the fundamentals in submitting a study to FDA, it's the process where data is formatted to follow a similar structure and organization, in order to facilitate the review process by regulatory agencies and improve the data quality [4,5]. Regarding SDTM datasets, they're standardized datasets for clinical trial raw data, with its development and maintenance being done by the Clinical Data Interchange Standards Consortium (CDISC). This format is used to ensure that clinical trial data is consistent, reliable, and easily shareable across different organizations and systems. Information collected from the study such as Subject Visits (SV), Adverse Events (AE) and Pharmacokinetics Concentrations (PC) is all represented in the SDTM datasets [4]. On the other hand, ADaM datasets are standardized for clinical trial analysis data. While ADaM is also developed and maintained by CDISC, this format is used to ensure that clinical trial data is consistently analyzed and reported, therefore, they are used to perform statistical analysis on the clinical study data [5]. Both SDTM and ADaM will be explained with further detail in a later chapter of this report (chapter 3).

Following this data standardization process is the data reporting process which is Tables, Figures and Listings (TFLs) generation. TFLs are one of the main components of a clinical study report (CSR), as they allow a presentation of the data and outcomes collected during the trial that is clear and concise, which in turn allows an easy interpretation of the results. TFLs also help to certify that the results and data collected are presented in a consistent and organized format, which facilitates the review and analysis of the results by regulatory agencies [6]. As with the standardized datasets, TFLs will be further explained in a later chapter of this report (chapters 2.2 and 4).

1.3. Objectives

The main goals of this report are to provide an overview of the best practices and guidelines in the programming of TFLs for regulatory submission of Phase I clinical trials, as well as to exemplify some specific procedures of the creation of TFLs using the SAS programming language.

2. State of the Art

2.1. What is Clinical Research?

Clinical research englobes all studies conducted in the medical sector with the goal of better understanding the complexity of human health and disease, with the primary goal of improving the methods used by doctors in the treatment and prevention of disorders.

Researchers learn about the physiology of the human body, the progression or regression of diseases over time, the pharmacokinetic or pharmacodynamic response of the body to potential treatments, the behaviors that promote health and prevent illness as well as those that do the opposite through the entire process of clinical research. It is worth mentioning that subjects who decide to participate in clinical research studies may or may not directly receive personal benefits as a result [7].

Studies that involve clinical research can be split in to two types, which are the clinical trials and observational studies [7].

Clinical trials are studies in which one or more interventions are given to participants in order to test their effects, hence why they are also known as interventional studies. Often, the interventions being studied are investigational, meaning they haven't been approved for use by doctors [7]. Clinical Trials are typically conducted in four development phases, each serving a different purpose and providing answers to distinct questions. During Phase I trials, a small group of individuals undergoes the initial testing of an experimental drug or treatment. This phase primarily focuses on assessing safety, establishing an appropriate dosage range, and identifying any potential side effects [8]. Regarding Phase II trials, a larger group of people is given the experimental drug or treatment to assess its effectiveness and safety [8]. On the other hand, in Phase III trials the experimental drug or treatment is administered to large groups of people to confirm its effectiveness, monitor side effects, compare it to commonly used treatments, and collect information that allows safe use of the intervention [8]. Finally, Phase IV trials are post-marketing

studies conducted after the regulatory agency's (FDA and EMA, for example) approval. They provide additional information on the risks, benefits, and best use of the treatment or drug [8].

Observational studies are a type of research where data is collected from individuals or already existing data sets. The information gathered is often related to the health, behaviors, or surroundings of participants. Unlike clinical trials, participants aren't given any intervention in observational studies, and when an intervention is being studied, normally participants are using it in their daily routines, which makes them useful in understanding how people's health is affected by their behavior. Some studies may use patient registries, which are collections of voluntarily provided data from patients. Researchers can quickly access data from hundreds or thousands of similar patients through these registries [7].

Clinical research in Portugal

Clinical research in Portugal is regulated by the Regulation EU No 536/2014 and by the Law n.º 21/2014 [9,10].

Regarding the Regulation EU No 536/2014, also known as the Clinical Trials Regulation, is a pharmaceutical legislation of the European Union (EU) which took effect on January 31st, 2022 [9]. The goal of this regulation is the creation of a environment that is favorable and appealing in the EU for large-scale clinical research, while making sure that subjects in clinical trials are ensured high standards of safety and public transparency [9]. The evaluation, authorization and supervision of clinical trials across the EU and European Economic Area (EEA) countries are streamlined by this regulation, which is responsible for their assessment and oversight [9]. Prior to the regulation, sponsors of clinical trials were required to submit applications to national competent authorities and ethics committees in each country, which made regulatory approval for conducting a clinical trial a time-consuming process [9]. Thankfully, sponsors can now use a single online platform known as the Clinical Trials Information System (CTIS) to submit one application for approval to run a clinical trial in multiple European countries, thus simplifying the process of conducting multinational trials [9]. Moreover, the Clinical Trials Information System allows EU Member States to evaluate and authorize such applications together, thereby increasing efficiency [9].

The regulation's goal is to encourage innovation and research in the EU and facilitate the performance of larger clinical trials in multiple EU Member States/EEA countries [9].

Other key advantages of the regulation include enhancing information-sharing and collective decision-making on clinical trials, improving transparency of clinical trial information, and ensuring that all participants in EU clinical trials have high standards of safety [9].

The Law n.º 21/2014, also known as the Clinical Research Law, which was later amended by laws n.º 73/2015 and n.º 49/2018, regulates any procedure related with clinical research done in Portugal [10].

This law has two main areas of focus: firstly, the regulation of clinical trials involving medicines for human use in accordance with Directive No. 2001/20/EC of the European Parliament and Council, which aims to ensure good clinical practices are followed in conducting such trials across member states [10]; secondly, the regulation of clinical investigation of medical devices, in line with the partial transposition of Directive No. 2007/47/EC. It must be noted though that due to the repeal of Directive No. 2001/20/EC by the Regulation No. 536/2014, a specific section of the aforementioned law concerning interventional clinical studies with medicines will be revoked [10].

In Portugal, the number of Clinical Trial Applications (CTA) submitted to INFARMED has shown a decrease between 2006 and 2011, followed by a slight increase between 2011 and 2019. However, in 2020 and forward, this number drastically increases, reaching a high of 230 submissions in 2022. The number of CTAs authorised by INFARMED as also followed a similar path, reaching a high of 155 authorisations in 2020. These changes can be explained by the demand of new vaccines and anti-viral treatments caused by the SARS-CoV-2 pandemic. Regarding the average time for authorisation, there was a slight decrease between 2007 and 2015, followed by a slight increase between 2015 and 2020. From 2020 forward, the average time has been rapidly increasing, reaching a high of 87 average calendar days (Figure 1) [11].

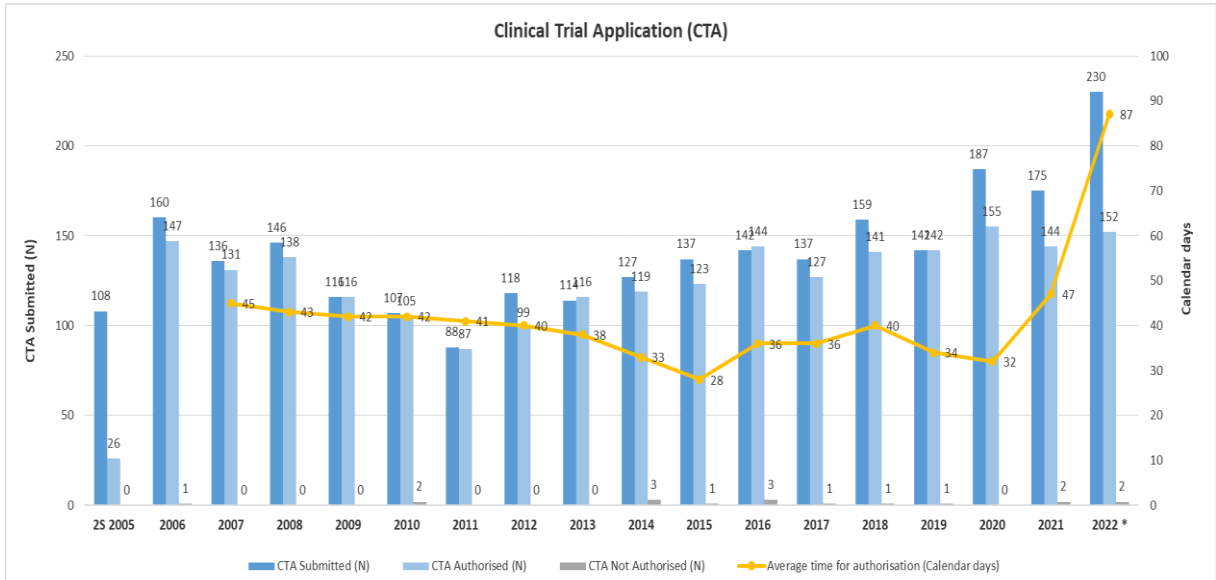


Figure 1 - Clinical trials in Portugal between the 2nd semester of 2005 and 2022. The dark blue bars represent the number of Clinical Trials submitted, the light blue bars represent the number of Clinical Trials authorized, and the gray bars represent the number of Clinical Trials not authorized. The orange line represents the average time in calendar days for authorization. Plot adapted from the 2005 to 4Q2022 Activity Indicators Statistics made available by INFARMED [11].

Most of the Clinical Trials performed in Portugal, from 2005 to 2022, have always been Phase III studies. However, it is interesting to notice that since 2012, there was an increase in Phase I studies performed in Portugal (Figure 2) [11].

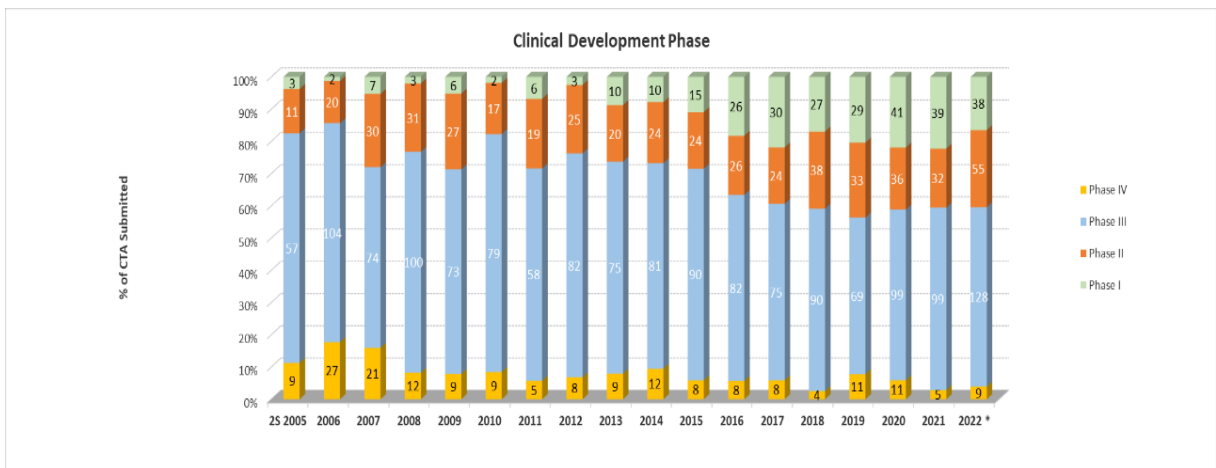


Figure 2 - Percentages of Clinical Trial applications submitted to INFARMED, divided by Clinical Development Phase. Phase I applications are indicated as green, Phase II applications are indicated as orange, Phase III applications are indicated as blue, and Phase IV applications are indicated as yellow. Plot adapted from the 2005 to 4Q2022 Activity Indicators Statistics made available by INFARMED [11].

Role of programming in Clinical Research

Programmers' ability to evaluate vast volumes of data and draw meaningful conclusions from it makes programming a crucial component of clinical research. One of the most popular tools in clinical programming is SAS (Statistical Analysis System), mainly due to its powerful statistical analysis capabilities, flexible data management features, and extensive documentation resources [12].

A notable strength of SAS is its ease of learning, as its syntax is very simple and therefore can be easily understandable even for those who don't have many experiences with programming. Additionally, SAS can handle large databases with ease, making it a preferred tool for big companies. The language also offers easy debugging through a comprehensible log window that clearly states errors [12].

SAS is well-known for its tested algorithms, as developers thoroughly test and analyze them before every version is released in a controlled environment. Furthermore, SAS offers excellent customer support through its closed source structure, which ensures that no external adulteration is possible [12].

Another advantage of SAS is its focus on data security, which is a critical feature for companies that handle confidential data. This language possesses a Graphical User Interface (GUI) that makes working with it simpler for users without programming experience, as well as a nice, structured output that is easily comprehensible. SAS is also a popular tool for job prospects in the analytics industry, as many professionals learn SAS as a prerequisite to entering the field [12].

While SAS offers many advantages, it also has some disadvantages compared to other programming languages like R. One of the major disadvantages of SAS is its cost, as it is a closed software that requires a proper license to use all its applications [12].

Another drawback of SAS is that it is not open source like R, meaning that its algorithms are not available for public research and are only accessible through a licensed version. In contrast, R has a greater availability for advanced graphics, with more descriptive plots, graphs, and diagrams [12].

Although SAS has drawbacks, it remains popular in the pharmaceutical sector, and regulatory bodies such as the FDA mandate that clinical trial data be analyzed with SAS. This requirement ensures that the analysis is conducted using a standardized and validated approach, providing confidence in the results and conclusions drawn from the study [13].

2.2. Regulatory landscape

International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use (ICH)

The International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use (ICH), which was founded in 1990, is crucial in facilitating discussions between regulatory agencies and the pharmaceutical industry about the scientific and technical facets of drug registration. By ensuring that the development and registration of safe, effective, and high-quality medicines is resource-efficient, ICH and has been continuously modified to meet the demands of the global environment in drug development. This has the purpose of promoting global harmonization. The ICH Guidelines, which were created with the active involvement of regulatory and industry professionals and are based on scientific consensus, have helped to achieve this goal [14].

Good Clinical Practice (GCP)

Good Clinical Practice (GCP) is an internationally recognized standard for the ethical and scientific aspects of trials that involve human subjects. This standard covers the design, implementation, documentation, and dissemination of trial results. While adhering to GCP, the protection of participants' rights, safety, and well-being is ensured, which reflects the principles that originated from the Declaration of Helsinki. Not only that, but GCP also enhances the credibility of clinical trial data [15].

ICH has developed the E6 (R2) Good clinical practice guideline, whose purpose is to establish a harmonized standard across the EU, Japan, and the United States, which enables regulatory authorities in these regions to mutually accept clinical data. When producing clinical trial data that is meant to be submitted to regulatory bodies, this guideline should be followed. Other clinical investigations that might have an effect on the security and welfare of human subjects may also use the guidelines specified in this guideline [15].

What are the main Regulatory Agencies involved?

As mentioned before, clinical studies must follow the requirements set by regulatory agencies in order to be submitted. In the case of BlueClinical, the main Regulatory Agencies involved in the studies are the U.S. Food and Drug Administration (FDA), the European Medicines Agency (EMA), the 'Comissão de Ética para a Investigação Clínica' (CEIC), and Infarmed.

Food and Drug Administration (FDA)

The FDA is the regulatory agency of the United States Department of Health and Human Services, having the responsibility to ensure the safety and efficacy of medications, including those used in clinical trials. The FDA accomplishes this by establishing a rigorous set of regulations that clinical trial sponsors must follow in order to conduct trials of new medications in the United States or to submit a trial [16].

One of the main roles of the FDA in clinical trials is the review and approval of the trial protocol before it can begin. The protocol must provide detailed information about the medication being tested, the trial population, the study design, and the endpoints that will be used to evaluate the medication's safety and effectiveness. The FDA carefully evaluates each protocol to ensure that it is scientifically sound and that it meets all of the agency's regulatory requirements [16]. If a major change to the protocol occurs, then the trial submission process must start over from scratch. This can only happen before the trial starts [17].

The FDA also monitors clinical trials as they are being conducted to ensure that they are being conducted in accordance with the approved protocol and that the medication is being administered safely. The agency has the power to stop a clinical trial if it determines that the medication is unsafe or if the trial is not being conducted in accordance with the approved protocol [17].

Once a clinical trial is completed, the FDA reviews the data to determine whether the medication is safe and effective for its intended use. The FDA also evaluates the benefit-risk profile of the medication to determine whether the potential benefits outweigh the potential risks. Based on this evaluation, the FDA may approve the medication for marketing in the United States, or it may require additional studies to be conducted [17].

FDA's regulatory requirements in clinical data programming

Regarding clinical data programming, FDA does have specific requirements which can be found in documents such as the Study Data Technical Conformance Guide [13], which is more specific to the process of submission to the FDA, and the ICH Electronic Common Technical Document (eCTD) Specification [6], which gives requirements regarding the TFLs' format.

Submissions to FDA

According with the module 5 of the eCTD, which is the module specific to the CSR, both SDTM and ADaM datasets must be submitted in the Transport Format (XPORT) format. This format was published by the SAS Institute as an open file format for exchanging study data. XPORT makes it easier for data to be translated into and out of various commonly used formats, without needing any vendor-specific programs. To ensure proper use, each transport file should contain just one dataset, and the dataset within the transport file must share the same name as the transport file itself. For instance, the SDTM dataset 'AE' would match to the transport file ae.xpt, 'PC' to pc.xpt, and the dataset 'ADAE' would correspond to adae.xpt. The source code used to create the datasets must also be submitted in single byte ASCII text format [13].

What is also required for submission to FDA is the Define.XML, which is a document that provides an exhaustive account of all aspects of data submissions, including definitions, algorithms, conventions, variable types, variable lengths, variable levels, and more. Additionally, if the Define.XML is deemed unprintable, a PDF version must be provided to the FDA [13].

At last, the source code used to create the TFLs included in the CSR must also be submitted in single byte ASCII text format as well [13].

Continuing with TFLs, some requirements regarding their format must be followed in order to submit them to the regulatory agencies. In the case of the FDA, the TFLs must follow the ICH eCTD Specification, more specifically the version 3.2.2. This document provides suggestions for file formats that are frequently employed in electronic submissions. However, depending on the guidance published in a particular region, alternative formats may also be acceptable [6].

These are the recommendations described in the eCTD specification regarding TFLs: Fonts such as Times New Roman, Arial, and Courier are acceptable to use [6]. Whenever feasible, narrative text should be typed in 12-point Times New Roman font, which is sufficiently large. In the case of TFLs, it is generally acceptable to use Times New Roman font sizes 9-10 or a similar size of other recommended fonts in tables. However, it is advisable to avoid using smaller font sizes. It must be noted that when selecting a font size for TFLs, it is essential to strike a balance between presenting enough information on a single page to enable reviewers to compare data while retaining a legible font size [6]. Regarding color, it is recommended to use black font in documents, while blue can be utilized for hypertext links. Light colors that are not suitable for grayscale printers should be avoided [6]. All TFLs should be correctly oriented, with portrait pages presented in portrait and landscape

pages presented in landscape [6], and their print area should fit on both A4 (210 x 297 mm) and Letter (8.5" x 11") paper. To avoid obscuring information when the reviewer subsequently prints and binds the pages for temporary use, a sufficient margin of at least 2.5 cm should be provided on the left side of each page. For landscape oriented TFLs, smaller margins of at least 2.0 cm at the top and 0.8 cm on both the left and right sides allow for more information to be displayed legibly on the page. Header and footer information may appear within these margins but should not be positioned too close to the page edge to prevent it from being lost during printing [6]. Regarding headers and footers, every TFL must feature a distinctive header or footer that briefly identifies the subject matter it pertains to [6]. Lastly, Pages of TFLs must be numbered consecutively with Arabic numerals. It is not appropriate to number pages using Roman numerals and leaving pages unnumbered is also not permissible [6].

Other important FDA guidelines

The FDA has also created other specific guidelines to ensure that data is collected, managed, and reported accurately, consistently, and with integrity. These guidelines include the 'Electronic Records; Electronic Signatures — Scope and Application (21 CFR Part 11)', which sets out the criteria for electronic records and electronic signatures that are equivalent to paper records and handwritten signatures, and as such, clinical data programming must comply with these requirements to ensure that electronic records and signatures are trustworthy, reliable, and equivalent to paper-based records [18], the 'Guidance for Industry: Computerized Systems Used in Clinical Investigations (2007)', which provides advice on the use of computerized systems in clinical studies, which include capture, management, and analysis of data, therefore emphasizing the need for validation, security, and audit trails to make sure that clinical data is accurate, reliable, and has integrity [19], the 'ICH Guidance for Industry: E9 Statistical Principles for Clinical Trials (1998)', which issues statistical principles for the design, conduct, and analysis of clinical trials, thus highlighting the need for appropriate statistical methods, data quality, and data consistency to ensure the validity and reliability of clinical trial results [3], the 'Guidance for Industry: Electronic Source Data in Clinical Investigations (2013)', which gives recommendations on the use of electronic source data (eSource) in clinical investigations, including data capture, management, and review, and as consequence, it points up the necessity for data accuracy, completeness, and consistency, as well as the need for a validated electronic system to collect and manage eSource data [20], and the 'Guidance for Industry: Providing Regulatory Submissions in Electronic Format — Standardized Study Data (2017)', which provides recommendations on the submission of standardized study data in electronic format,

bringing therefore attention to the requirements for data quality, consistency, and standardization to facilitate the review and analysis of clinical data by regulatory authorities [21].

European Medicines Agency (EMA)

The EMA is the regulatory agency of the European Union responsible for the evaluation, supervision, and approval of medicines for human use.

The role of EMA in clinical trials is to ensure the safety, effectiveness and high quality of the medicines being developed. The agency assesses the results of clinical trials submitted by pharmaceutical companies and determines if the benefits of the medicine do outweigh the risks for patients or not [22,23].

As a requirement by EMA, all clinical trials conducted within the EU must adhere to its strict ethical and scientific standards, which are summarized in the EU Clinical Trials Regulation. This regulation sets out requirements when it comes to trial design, conduct, and reporting and is applied to all clinical trials that are conducted in the EU, regardless of the location of the sponsor or investigator [22,23].

EMA also guides pharmaceutical companies on the design and conduct of clinical trials, as well as on the format and content of regulatory submissions. The agency works closely with national regulatory authorities in the EU member states to make sure that the evaluation of medicines is consistent [22].

One example of the role of EMA in clinical trials can be seen in the approval of the COVID-19 vaccines. The agency reviewed the results of the clinical trials for the vaccines developed by Pfizer-BioNTech, Moderna, AstraZeneca, and Johnson & Johnson, and granted conditional marketing authorizations based on their safety and efficacy profiles [24].

EMA's regulatory requirements in clinical data programming

Regarding clinical data programming, EMA has provided guidance for the regulatory requirements. Some of the guidance is found in ICH guidelines, which include the ICH E6(R2) guideline for GCP [15] and the ICH M4 (R4) on common technical documents (CTD) [26]. This latter document specifies that when preparing TFLs, it is crucial to use margins that enable the document to be printed on A4 paper. It also tells that the fonts used should be of a size and style that ensures legibility, even after photocopying. A font size of 12-point Times New Roman is recommended. It is also important to number every page in accordance with the granularity of the document [26].

Another relevant ICH document is the M2 EWG eCTD Specification, which was mentioned before in chapter 2.2. A noteworthy module mentioned in this document is the module 5, which describes the organization of CSRs, and therefore, the TFLs to be submitted and how [6].

EMA has also developed other guidelines besides the ICH ones, which include the EMA Guideline on Good Pharmacovigilance Practices (GVP), that provides recommendations for the collection, management, and reporting of adverse drug reactions (ADRs) during clinical trials and therefore outlines the requirements for clinical data programming related to pharmacovigilance [26], the EMA Guideline on Data Monitoring Committees (DMCs), which gives advice for the establishment and operation of DMCs, which oversee the safety and efficacy of clinical trials, thus summarizing the demands for clinical data programming connected to DMC activities [27], and the EMA Guideline on Electronic Data Capture (EDC), which issues guidance for the use of EDC in clinical trials by resuming the necessary aspects of clinical data programming associated to the use of EDC systems [28].

‘Comissão de Ética para a Investigação Clínica’ (CEIC)

Comissão de Ética para a Investigação Clínica (CEIC) is a regulatory body in Portugal that oversees ethical aspects of clinical research. Among its responsibilities is the reviewing and approving of protocols for clinical trials and other research studies that involve human subjects, in order to allow them to be performed so that the safety, rights, and welfare of the subjects involved can be protected [29].

The CEIC was founded in Portugal in 2005 as part of the Portuguese National Health System, therefore under the authority of the Portuguese Ministry of Health. This regulatory body embodies a multidisciplinary team of experts, from physicians to ethicists and legal experts, that reviews proposals for clinical trials so that they meet ethical and regulatory standards [29].

According to the Portuguese legislation, CEIC is a mandatory entity for all clinical research projects that involve human beings, their tissues, or data and that are performed in Portugal. According to the National Regulation of Clinical Trials (Law nº 21/2014), for a clinical trial to proceed it must get the approval of the CEIC, and researchers must obtain informed consent from participants prior to their participation in the study [10,30].

The CEIC also plays a crucial role in ensuring that clinical research in Portugal complies with international ethical standards, including the ICH guidelines, the Declaration of Helsinki, and the GCP guidelines [25,30].

CEIC role on Clinical Data Programming

While the CEIC is responsible for reviewing and approving research protocols, just as mentioned before, it does not however have a direct role in clinical data programming.

However, the CEIC does ensure that ethical principles related to data privacy and confidentiality are respected, as the handling and processing of clinical data is a critical aspect of protecting research participants' safety, rights, and welfare. In this regard, the CEIC may require that the researchers establish data protection measures, such as pseudonymization, to ensure that personal information is kept confidential [30].

In addition, CEIC may need data management and analysis plans that agree with the study protocol, to make sure the data is exact, thorough, and pertinent to the study. Furthermore, CEIC may review informed consent procedures to ensure that subjects are informed about data collection, its use, and who will have access to it [10].

It is worth noting that in Portugal, clinical data programming is subject to the provisions of the General Data Protection Regulation (GDPR) and the Portuguese Data Protection Law (Lei de Proteção de Dados Pessoais). These regulations that have been implemented are incredibly strict when it comes to the collecting, processing, and storing of personal data, including clinical data, and state that the researchers must get the participants' full consent and understanding before data processing occurs [30,31].

Infarmed

INFARMED - Autoridade Nacional do Medicamento e Produtos de Saúde, I.P., or Infarmed for short, is the National Authority of Medicines and Health Products in Portugal. INFARMED's primary objective is to oversee and enforce regulations on medications, medical devices, and personal care products while upholding the utmost standards of public health preservation. Their purpose is to guarantee that the access to medications, medical devices, cosmetics, and hygiene products to healthcare personnel and the public is safe, efficient, and reliable. Infarmed also regulates clinical trials and guides the use of medicines and other health products [32].

Founded in 1993, just like CEIC, Infarmed is part of the Portuguese National Health System, and regulated by the Portuguese Ministry of Health. Its mission is to safeguard and foster public health by guaranteeing that medicines and other health products are safe, effective, and of top-notch quality [32].

Infarmed is responsible for several important tasks related to health policy. These tasks include contributing to the development of policies for medicines and health products, such as medical devices and hygiene products. Infarmed is also responsible for regulating, evaluating, and supervising research, production, distribution, commercialization, and use of medicines and health products. This includes ensuring compliance with clinical trial authorization rules and good clinical practices. Infarmed also monitors the consumption and use of these products and promotes access to information for health professionals and consumers [32].

Additionally, Infarmed promotes and supports research in pharmaceutical science and technology, biotechnology, pharmacology, pharmacoeconomics, and pharmacoepidemiology, in collaboration with universities and research institutions. Infarmed ensures appropriate integration within the European Union system for evaluation and supervision of medicinal products, including liaison with the European Medicines Agency and other institutions. It also participates in the European Union network of authorities for medicinal products and official quality control laboratories for medicinal products [32].

Infarmed assures the State's international obligations within its scope, including those within the European Union, Council of Europe, European Pharmacopoeia Commission, and United Nations Organization. It also develops national and international cooperation activities and manages the National Health Technology Assessment System (SiNATS) [32].

Infarmed role on Clinical Data Programming

Infarmed is not directly involved in clinical data programming in Portugal. However, as the national regulatory agency responsible for overseeing clinical trials, Infarmed has a crucial role in ensuring that the handling and processing of clinical data comply with ethical and legal standards.

Infarmed ensures that the conduct of clinical trials in Portugal is carried out under international standards and applicable laws and regulations, including those related to data protection and privacy. The regulatory requirements for the handling and processing of clinical data are specified in Portuguese Law no. 21/2014, which regulates clinical research in Portugal, and in the General Data Protection Regulation (GDPR), which is applicable across the European Union [10,30]

In order to fulfill these regulations and to protect the privacy rights of the subjects involved, as well as its confidentiality, integrity, and availability, clinical data must be processed. Infarmed's role in this regard is to ensure that the clinical trial protocol and the data management plan are

consistent with the applicable regulatory requirements and that the sponsor or investigator has obtained the participants' informed consent for data processing [10].

Moreover, Informed also requires that clinical trial data are collected, recorded, and reported in a way that ensures their accuracy, completeness, and reliability. This includes ensuring that the data are consistent with the clinical trial protocol and that the data management plan specifies the procedures for handling, processing, and storing the data [10].

2.3. Clinical Study Documents to be considered for TFLs

There are certain documents that must be considered regarding TFLs, with the main ones being the Statistical Analysis Plan (SAP), which will the structure and organization that each TFL must follow, and the CSR, the document where TFLs will be integrated.

Statistical Analysis Plan (SAP)

The SAP is a document that outlines the procedures and methods that will be used to analyze data collected in a research study or clinical trial. It serves as a blueprint for the statistical analysis and provides a clear description of the statistical methods that will be employed to test the research hypotheses or objectives [18].

A well-written SAP should explain the study design, statistical methods, variables, and assumptions. The SAP should also describe the data management and quality control procedures that will be used to ensure the accuracy and reliability of the data [18].

The SAP is typically developed early in the study planning process, often before data collection has begun, to ensure that the data collected will be sufficient to answer the research questions. The SAP may be updated throughout the study to reflect changes in the study design or data collection procedures, but any changes should be documented and justified [18].

Regarding TFLs, the SAP also describes how the study data and analysis results will be summarized and reported in the CSR, by explaining the structure and organization the TFLs must have [18].

Clinical Study Report (CSR)

A Clinical Study Report (CSR) is a document that details a clinical trial's design, methods, findings, and conclusions, being usually written by the sponsor. [33].

The CSR covers background, objectives, selection methods, treatment to be tested, safety, effectiveness, and results, including side effects. Additionally, information such as statistical analysis of data and summaries of individual participant data can be included in the CSR. [33]

The CSR should include data presented at numerous levels of detail. It is recommended to include overall summary figures and tables in the text to highlight key points related to demographic, efficacy, and safety variables. Additional summary figures and tables for demographic, efficacy, and safety variables should be provided in the section 14 of the CSR. For specified groups of patients, individual patient data should be presented as listings in CSR Appendix 16.2. Lastly, the CSR Appendix 16.4 should contain all individual patient data, with archival listings requested exclusively for the US [33].

3. Data Standardization

Data Standardization is the process in which data is formatted in order to follow a standardized structure and organization. In clinical studies, data is mostly standardized according with the Study Data Tabulation Model (SDTM) and Analysis Data Model (ADaM) standards, both developed by the Clinical Data Interchange Standards Consortium (CDISC), since it is a requirement by many regulatory agencies, such as the FDA [13]. This is mainly due to the benefits that these standards offer, which include fostered efficiency, complete traceability, enhanced innovation, improved data quality, facilitated data sharing, reduced costs, increased predictability, and streamlined processes [34]. Just as a quick reminder, the created standardized datasets must also be in the xpt. format in order to be submitted, just as mentioned in chapter 2.2.

3.1. SDTM Datasets

Regarding the SDTM datasets, these are used to represent data collected during a clinical trial in a standardized format. In the process of developing these datasets, two distinct documents must be kept in mind, the SDTM Specification [35], and the SDTM Implementation Guide (SDTMIG) [36], which defines respectively the main guidelines and the implementation details that need to be strictly followed during the process.

In SDTM datasets, each observation is represented by a series of named variables that can be classified based on their role. There are five major roles for SDTM variables: identifier variables that identify the study, subject, domain, and sequence number; topic variables that specify the focus of

the observation; timing variables that describe the timing of an observation; qualifier variables that provide additional descriptive text or numeric values; and rule variables that express algorithms or executable methods for defining start, end, or looping conditions in the Trial Design model [35,36].

Regarding qualifier variables, these can be classified into five sub-classes based on their purpose. Grouping Qualifiers are used to group together a collection of observations within the same domain. Result Qualifiers describe the specific results associated with the topic variable in a Findings dataset and answer the question raised by the topic variable. Synonym Qualifiers specify an alternative name for a particular variable in an observation. Record Qualifiers define additional attributes of the observation record as a whole. Variable Qualifiers are used to further modify or describe a specific variable within an observation and are only meaningful in the context of the variable they qualify. Examples of these sub-classes include --CAT and --SCAT as Grouping Qualifiers, -ORRES and --STRESC as Result Qualifiers, --MODIFY and --DECOD as Synonym Qualifiers, --REASND and --POS as Record Qualifiers, and --ORRESU and --DOSU as Variable Qualifiers [35,36].

In SDTM, variables are also categorized into three core designations: Required, Expected, and Permissible. A Required variable is necessary to identify a data record or to make it meaningful. It is mandatory and cannot have null values for any record. An Expected variable is necessary to provide context within a specific domain and may have some null values, but not for every record. If an expected variable is not included in the study, a null column must still be present in the dataset. A Permissible variable is optional but should be used whenever relevant. All variables within the Identifier, Timing, and general observation classes are permissible unless restricted by the implementation guide or domain assumptions [35,36].

Listed below are examples of some of the most notable SDTM datasets, especially in the making of Tables, Figures and Listings (TFLs). It must be kept in mind that these datasets, and the ones not listed, are usually accompanied with other datasets when analyzed, in order to get a better comprehension of the study's results and subjects. Regarding these specific examples, it must be emphasized that the data used was a simulated which was created to describe a two treatment two sequence, two periods of a safety randomized study that compares the fictional drugs Drug A, which is the test drug, and Drug B, which is the reference. This study will have sequence 1 and sequence 2 has the two study sequences and for the data simulation, five subjects plus 1 screening failure and a subject that wasn't assigned to the study will be included. The software used to create the following SDTM datasets was the SAS Enterprise Guide 8.3, which runs on SAS 9.4.

Subject Visits (SV)

The SV dataset is a vital component of the Study Data Tabulation Model (SDTM) utilized for analyzing clinical trial data. It records details about the actual visits made by a study participant throughout the trial. This dataset plays a significant role in organizing and examining information linked to a subject's visits, including assessments, interventions, and follow-up [35,36].

The importance of the SV dataset is multi-fold. Firstly, it enables the organization and analysis of a subject's visit-related data, providing insights into the timing and frequency of assessments, interventions, and follow-up. Secondly, it helps identify any missing or incomplete data that may be crucial in interpreting study results. Finally, the SV dataset is employed in safety analyses, where the severity and frequency of adverse events are compared between treatment groups [36,37].

The SV dataset typically includes the following variables [35,36]:

- STUDYID: A unique identifier for the study.
- DOMAIN: The domain abbreviation (SV).
- USUBJID: A unique identifier for each study subject (for this specific study the rule applied was STUDYID + “_” + Randomization number).
- VISITNUM: The visit number for the subject.
- VISIT: The name of the visit as defined in the study protocol.
- VISITDY: The planned study day on which the visit was scheduled to occur.
- SVSTDTC: The actual date and time of the visit.
- SVENDTC: The actual end date and time of the visit.

STUDYID	DOMAIN	USUBJID	VISITNUM	VISIT	VISITDY	EPOCH	SVSTDT	SVENDT	SVSTDY	SVENDY
Bioinf-tese-01	SV	Bioinf-tese-01_01	1.0	Screening	-28	SCREENING	2012-12-26	2013-01-18T17:00	-24	-1
Bioinf-tese-01	SV	Bioinf-tese-01_01	2.0	Period 1: Admission	-1	PERIOD 1	2013-01-18T17:00	2013-01-19T07:00	-1	1
Bioinf-tese-01	SV	Bioinf-tese-01_01	3.0	Period 1: Confinement	1	PERIOD 1	2013-01-19T07:00	2013-01-20T08:00	1	2
Bioinf-tese-01	SV	Bioinf-tese-01_01	4.0	Period 1: Ambulatory	2	PERIOD 1	2013-01-20T08:00	2013-01-20T19:00	2	2
Bioinf-tese-01	SV	Bioinf-tese-01_01	5.0	Period 2: Admission	7	PERIOD 2	2013-01-25T17:00	2013-01-26T07:00	7	8
Bioinf-tese-01	SV	Bioinf-tese-01_01	6.0	Period 2: Confinement	8	PERIOD 2	2013-01-26T07:00	2013-01-27T08:00	8	9
Bioinf-tese-01	SV	Bioinf-tese-01_01	7.0	Period 2: Ambulatory	9	PERIOD 2	2013-01-27T08:00	2013-01-27T19:00	9	9
Bioinf-tese-01	SV	Bioinf-tese-01_01	8.0	End of Study	9	END OF STUDY	2013-01-27T19:35	2013-01-27T19:35	9	9
Bioinf-tese-01	SV	Bioinf-tese-01_02	1.0	Screening	-28	SCREENING	2012-12-26	2013-01-18T18:00	-24	-1
Bioinf-tese-01	SV	Bioinf-tese-01_02	2.0	Period 1: Admission	-1	PERIOD 1	2013-01-18T18:00	2013-01-19T07:00	-1	1
Bioinf-tese-01	SV	Bioinf-tese-01_02	3.0	Period 1: Confinement	1	PERIOD 1	2013-01-19T07:00	2013-01-20T08:00	1	2
Bioinf-tese-01	SV	Bioinf-tese-01_02	4.0	Period 1: Ambulatory	2	PERIOD 1	2013-01-20T08:00	2013-01-20T19:00	2	2
Bioinf-tese-01	SV	Bioinf-tese-01_02	5.0	Period 2: Admission	7	PERIOD 2	2013-01-25T18:00	2013-01-26T07:00	7	8
Bioinf-tese-01	SV	Bioinf-tese-01_02	6.0	Period 2: Confinement	8	PERIOD 2	2013-01-26T07:00	2013-01-27T08:00	8	9
Bioinf-tese-01	SV	Bioinf-tese-01_02	7.0	Period 2: Ambulatory	9	PERIOD 2	2013-01-27T08:00	2013-01-27T19:00	9	9
Bioinf-tese-01	SV	Bioinf-tese-01_02	8.0	End of Study	9	END OF STUDY	2013-01-27T19:40	2013-01-27T19:40	9	9
Bioinf-tese-01	SV	Bioinf-tese-01_03	1.0	Screening	-28	SCREENING	2012-12-26	2013-01-18T18:00	-24	-1
Bioinf-tese-01	SV	Bioinf-tese-01_03	2.0	Period 1: Admission	-1	PERIOD 1	2013-01-18T18:00	2013-01-19T07:00	-1	1
Bioinf-tese-01	SV	Bioinf-tese-01_03	3.0	Period 1: Confinement	1	PERIOD 1	2013-01-19T07:00	2013-01-20T08:00	1	2
Bioinf-tese-01	SV	Bioinf-tese-01_03	4.0	Period 1: Ambulatory	2	PERIOD 1	2013-01-20T08:00	2013-01-20T19:00	2	2
Bioinf-tese-01	SV	Bioinf-tese-01_03	8.0	End of Study	9	END OF STUDY	2013-01-23T13:41	2013-01-27T14:21	5	9

Figure 3 - Simplified Example of a SV dataset.

Regarding the SV dataset represented in Figure 3, we notice that the subject Bioinf-tese-01_03 has ended the study prior to the Period 2 visits. This event will be further detailed in the next two datasets.

Demographics (DM)

The SDTM (Study Data Tabulation Model) utilizes the DM dataset to gather vital information about the demographic attributes of study participants in clinical trials. This information can assist researchers in comprehending the patient population being studied and in examining the trial result [35,36].

The SDTM (Study Data Tabulation Model) utilizes the DM dataset to gather vital information about the demographic attributes of study participants in clinical trials. This information can assist researchers in comprehending the patient population being studied and in examining the trial result [35,36].

The DM dataset holds a crucial position for various reasons. Firstly, individual study subjects can be identified and monitored during the course of the trial. Not only that, but the influence of demographic factors on the study outcomes can also be examined. For instance, different age groups, genders, or races can be used to compare the results of the trial. Finally, the DM dataset is commonly employed in safety evaluations, as it helps in identifying any demographic factors that may contribute to higher risks of adverse events or other safety concerns [35,36].

The DM dataset usually includes the following variables [35,36]:

- STUDYID: A unique identifier for the study.
- DOMAIN: The domain abbreviation (DM).
- USUBJID: A unique identifier for each study subject.
- SUBJID: A subject identifier that may be used by the site.
- SITEID: A site identifier that may be used by the sponsor.
- RFSTDTC: Subject Reference Start Date/Time.
- BRTHDTC: The date of birth of the study subject.
- DTHDTC: The date of death of the study subject.
- AGE: The age of the study subject.
- SEX: The sex of the study subject.
- RACE: The race of the study subject.
- ARM: The study arm to which the subject was assigned according to plan.
- ACTARM: The study arm that the subject did actually follow.
- COUNTRY: The country in which the study was conducted.

STUDYID	DOMAIN	USUBJID	SUBJID	RFSTDTC	RFICDTC	DTHDTC	SITEID	SEX	RACE	ARM	ACTARM	COUNTRY
Bioinf-tese-01	DM	Bioinf-tese-01_01	1	2013-01-19T07:30:00	2012-12-26		BlueClinical Phase I	F	ASIAN	Test-Reference	Test-Reference	PRT
Bioinf-tese-01	DM	Bioinf-tese-01_02	2	2013-01-19T07:33	2012-12-26		BlueClinical Phase I	M	MULTIPLE	Reference-Test	Reference-Test	PRT
Bioinf-tese-01	DM	Bioinf-tese-01_03	3	2013-01-19T07:36:10	2012-12-26		BlueClinical Phase I	F	WHITE	Test-Reference	Unplanned Treatment	PRT
Bioinf-tese-01	DM	Bioinf-tese-01_04	4	2013-01-19T07:39	2012-12-27		BlueClinical Phase I	M	BLACK OR AFRICAN AMERICAN	Reference-Test	Reference-Test	PRT
Bioinf-tese-01	DM	Bioinf-tese-01_05	5	2013-01-19T07:42:00	2012-12-27		BlueClinical Phase I	F	WHITE	Test-Reference	Test-Reference	PRT
Bioinf-tese-01	DM	SCR_06	6		2012-12-26		BlueClinical Phase I	M	WHITE	Not Assigned	Not Assigned	PRT
Bioinf-tese-01	DM	SCR_07	7		2012-12-28		BlueClinical Phase I	M	MULTIPLE	Screen Failure	Screen Failure	PRT

Figure 4 - Simplified Example of a DM dataset.

Regarding the DM dataset represented in Figure 4, we notice that the subject Bioinf-tese-01_03 did not perform the planned sequence of treatments that was attributed to him. Linking this observation with the one mentioned in the SV dataset, we can conclude that the subject only took one of the two planned treatments, the drug A to be precise, but since he was discontinued in this study before doing any Period 2 visits, he didn't take the drug B. The reason for this discontinuation is due to an adverse event that occurred on the subject, which will be described later in the next dataset.

Adverse Events (AE)

In the SDTM (Study Data Tabulation Model), the Adverse Events (AE) dataset holds significant importance as it records the occurrences of any undesirable events or side effects encountered by participants in the trial [35,36].

This dataset serves multiple purposes. Firstly, it facilitates monitoring and evaluation of any AE that arise during the study, including their severity and outcome. This information aids researchers in identifying safety issues associated with the intervention or drug under investigation. Secondly, the AE dataset is utilized in safety analyses to compare the frequency and severity of adverse events across treatment groups. Finally, the AE dataset is valuable for post-marketing surveillance to monitor the safety of the drug after its approval and release to the public [35,36].

The AE dataset generally includes the following variables [35,36]:

- STUDYID: A unique identifier for the study.
- DOMAIN: The domain abbreviation (AE).
- USUBJID: A unique identifier for each study subject.
- AESEQ: A sequence number that identifies each adverse event for a subject.
- AETERM: The adverse event term, which describes the specific symptom or condition experienced by the study subject.
- AESTDTC: The event onset date, which is the date on which the adverse event was first observed or reported.
- AEENDTC: The end date of the adverse event, if applicable.
- AESER: The severity of the adverse event, which is often graded on a scale from mild to severe.
- AEOUT: The outcome of the adverse event, such as whether it resolved, improved, or resulted in death.
- AEBODSYS: The body system affected by the adverse event.
- AEREL: The relationship of the adverse event to the study drug or other interventions.
- AEACN: The action taken with study treatment.
- AEDECOD: The MedDRA (Medical Dictionary for Regulatory Activities) code for the adverse event.

STUDYID	DOMAIN	USUBJID	AESEQ	AETERM	AESEV	AESER	AEACN	AEREL	AEOUT	AESTDTC	AEENDTC
Bioinf-tese-01	AE	Bioinf-tese-01_01	1	Somnolence	MODERATE	N	DOSE NOT CHANGED	REASONABLY POSSIBLE	RECOVERED/RESOLVED	2013-01-19T07:50	2013-01-19T15:30
Bioinf-tese-01	AE	Bioinf-tese-01_01	2	Nausea	MILD	N	DOSE NOT CHANGED	REASONABLY POSSIBLE	RECOVERED/RESOLVED	2013-01-19T10:00	2013-01-19T15:00
Bioinf-tese-01	AE	Bioinf-tese-01_02	1	Somnolence	MODERATE	N	DOSE NOT CHANGED	REASONABLY POSSIBLE	RECOVERED/RESOLVED	2013-01-19T09:00	2013-01-19T15:30
Bioinf-tese-01	AE	Bioinf-tese-01_02	2	Soft stools	MILD	N	NOT APPLICABLE	REASONABLY POSSIBLE	RECOVERED/RESOLVED	2013-01-19T11:00	2013-01-19T11:00
Bioinf-tese-01	AE	Bioinf-tese-01_03	1	Somnolence	MILD	N	DOSE NOT CHANGED	REASONABLY POSSIBLE	RECOVERED/RESOLVED	2013-01-19T08:30	2013-01-19T16:00
Bioinf-tese-01	AE	Bioinf-tese-01_03	2	Flatulence	MILD	N	DOSE NOT CHANGED	REASONABLY POSSIBLE	RECOVERED/RESOLVED	2013-01-19T11:00	2013-01-19T20:00
Bioinf-tese-01	AE	Bioinf-tese-01_03	3	Flu like syndrome	MILD	N	DRUG WITHDRAWN	REASONABLY POSSIBLE	RECOVERED/RESOLVED	2013-01-22T16:00	2013-01-23T07:00
Bioinf-tese-01	AE	Bioinf-tese-01_04	1	Rhinopharyngitis	MILD	N	NOT APPLICABLE	NOT REASONABLY POSSIBLE	RECOVERED/RESOLVED	2013-01-03T06:00	2013-01-03T18:00
Bioinf-tese-01	AE	Bioinf-tese-01_04	2	Somnolence	MILD	N	DOSE NOT CHANGED	REASONABLY POSSIBLE	RECOVERED/RESOLVED	2013-01-19T08:30	2013-01-19T16:00
Bioinf-tese-01	AE	Bioinf-tese-01_04	3	Constipation	MILD	N	DOSE NOT CHANGED	REASONABLY POSSIBLE	RECOVERED/RESOLVED	2013-01-26T15:00	2013-01-26T06:00
Bioinf-tese-01	AE	Bioinf-tese-01_05	1	Dry cough	MILD	N	NOT APPLICABLE	NOT REASONABLY POSSIBLE	RECOVERED/RESOLVED	2013-01-16T09:00	2013-01-19T12:00
Bioinf-tese-01	AE	Bioinf-tese-01_05	2	Somnolence	MILD	N	DOSE NOT CHANGED	REASONABLY POSSIBLE	RECOVERED/RESOLVED	2013-01-19T08:30	2013-01-19T18:00
Bioinf-tese-01	AE	Bioinf-tese-01_05	3	Frontal headache	MILD	N	DOSE NOT CHANGED	REASONABLY POSSIBLE	RECOVERED/RESOLVED	2013-01-19T16:00	2013-01-20T06:00
Bioinf-tese-01	AE	Bioinf-tese-01_05	4	Somnolence	MODERATE	N	NOT APPLICABLE	REASONABLY POSSIBLE	RECOVERED/RESOLVED	2013-01-26T08:45	2013-01-26T19:00

Figure 5 - Simplified Example of a AE dataset.

In the AE dataset exemplified in the figure 3, all the AEs that occurred in the subjects can be observed. Regarding the subject Bioinf-tese-01_03, it is noticed that one of the AEs that occurred on her was a flu like syndrome, which as a response, it was decided to withdraw the treatment on her. In other words, this AE is the reason why this subject discontinued this study.

Exposure (EX)

The Study Data Tabulation Model (SDTM) includes a crucial element called the EX dataset. This dataset records vital information regarding the exposure of trial participants to the investigational drug or intervention being tested [35,36].

The EX dataset holds significant importance for several reasons. Firstly, exposure of study subjects to the investigational drug or intervention can be monitored and evaluated, aiding in determining its efficacy and identifying any adverse events or safety concerns associated with its use. Secondly, the EX dataset is utilized in pharmacokinetic and pharmacodynamic analyses to investigate the correlation between drug exposure and its impact on the body. Lastly, the EX dataset can be employed in post-marketing surveillance to track patient exposure to the drug or intervention after it has received approval and been released to the general population [35,36].

The EX dataset commonly includes the following variables [35,36]:

- STUDYID: A unique identifier for the study.
- DOMAIN: The domain abbreviation (EX).
- USUBJID: A unique identifier for each study subject.
- EXSEQ: A sequence number that identifies each exposure for a subject.
- EXTRT: The actual treatment received by the subject.
- EXDOSE: The dose of the investigational drug or intervention.
- EXDOSU: The dose unit of the investigational drug or intervention.
- EXDOSFRM: The form of the investigational drug or intervention.

- EXSTDTC: The start date of the exposure.
- EXENDTC: The end date of the exposure.

STUDYID	DOMAIN	USUBJID	EXSEQ	EXTRT	EXDOSE	EXDOSU	EXDOSFRM	EPOCH	EXSTDTC	EXENDTC
Bioinf-tese-01	EX	Bioinf-tese-01_01	1	Drug A	35 mg		TABLET, ORALLY DISINTEGRATING	PERIOD 1	2013-01-19T07:30:00	2013-01-19T07:38:16
Bioinf-tese-01	EX	Bioinf-tese-01_01	2	Drug B	35 mg		TABLET	PERIOD 2	2013-01-26T07:30	2013-01-26T07:30
Bioinf-tese-01	EX	Bioinf-tese-01_02	2	Drug A	35 mg		TABLET, ORALLY DISINTEGRATING	PERIOD 2	2013-01-26T07:33:00	2013-01-26T07:37:00
Bioinf-tese-01	EX	Bioinf-tese-01_02	1	Drug B	35 mg		TABLET	PERIOD 1	2013-01-19T07:33	2013-01-19T07:33
Bioinf-tese-01	EX	Bioinf-tese-01_03	1	Drug A	35 mg		TABLET, ORALLY DISINTEGRATING	PERIOD 1	2013-01-19T07:36:10	2013-01-19T07:37:06
Bioinf-tese-01	EX	Bioinf-tese-01_04	2	Drug A	35 mg		TABLET, ORALLY DISINTEGRATING	PERIOD 2	2013-01-26T07:39:00	2013-01-26T07:40:34
Bioinf-tese-01	EX	Bioinf-tese-01_04	1	Drug B	35 mg		TABLET	PERIOD 1	2013-01-19T07:39	2013-01-19T07:39
Bioinf-tese-01	EX	Bioinf-tese-01_05	1	Drug A	35 mg		TABLET, ORALLY DISINTEGRATING	PERIOD 1	2013-01-19T07:42:00	2013-01-19T07:42:30
Bioinf-tese-01	EX	Bioinf-tese-01_05	2	Drug B	35 mg		TABLET	PERIOD 2	2013-01-26T07:42	2013-01-26T07:42

Figure 6 - Simplified Example of a EX dataset.

In the EX dataset represented in Figure 6, it is noticed that all study subjects have a record for each of the study treatments, apart from subject Bioinf-tese-01_03, which only has a record for Drug A but not for Drug B. This absence isn't surprising, considering that this subject was absent during the Period 2 of the study and was supposed to take the Drug B during this period, as observed on the datasets SV and DM respectively.

Pharmacokinetics Concentrations (PC)

The SDTM's PC dataset records data on the concentration of a drug or its metabolites in biological samples collected from study participants during clinical trials [35,36].

There are several reasons why the PC dataset is significant. Firstly, it enables the tracking and analysis of drug pharmacokinetics, aiding in the determination of optimal dosing schedules and identifying any safety concerns. Secondly, the PC dataset is utilized in pharmacokinetic analyses to investigate the connection between drug exposure and its effects on the body. Finally, the PC dataset can be leveraged in post-marketing surveillance to monitor drug concentration in patients after its approval and release to the general public [35,36].

The PC dataset normally includes the following variables [35,36]:

- STUDYID: A unique identifier for the study.
- DOMAIN: The domain abbreviation (PC).
- USUBJID: A unique identifier for each study subject.
- PCSEQ: A sequence number that identifies each concentration measurement for a subject.
- PCTESTCD: The code for the analytical method used to measure the drug concentration.
- PCTEST: The name of the analytical method used to measure the drug concentration.
- PCORRES: The numerical result of the concentration measurement.

- PCORRESU: The unit of measure for the concentration result.
- PCDTC: The date/time at which the concentration measurement was taken.
- PCSPEC: The type of specimen used for the concentration measurement.
- PCNAM: The analytical laboratory used for the concentration measurement.

STUDYID	DOMAIN	USUBJID	PCSEQ	PCTESTCD	PCTEST	PCORRES	PCORRESU	EPOCH	PCDTC	PCLTM
Bioinf-tese-01	PC	Bioinf-tese-01_01	1	A	ANALYTE A	BLQ	ng/mL	PERIOD 1	2013-01-19T07:15	-PT30M
Bioinf-tese-01	PC	Bioinf-tese-01_01	2	A	ANALYTE A	180.4	ng/mL	PERIOD 1	2013-01-19T08:30	PT1H
Bioinf-tese-01	PC	Bioinf-tese-01_01	3	A	ANALYTE A	436.6	ng/mL	PERIOD 1	2013-01-19T09:30	PT2H
Bioinf-tese-01	PC	Bioinf-tese-01_01	4	A	ANALYTE A	459.6	ng/mL	PERIOD 1	2013-01-19T10:30	PT3H
Bioinf-tese-01	PC	Bioinf-tese-01_01	5	A	ANALYTE A	428.5	ng/mL	PERIOD 1	2013-01-19T11:30	PT4H
Bioinf-tese-01	PC	Bioinf-tese-01_01	6	A	ANALYTE A	267.9	ng/mL	PERIOD 1	2013-01-19T13:30	PT6H
Bioinf-tese-01	PC	Bioinf-tese-01_01	7	A	ANALYTE A	181.2	ng/mL	PERIOD 1	2013-01-19T15:30	PT8H
Bioinf-tese-01	PC	Bioinf-tese-01_01	8	A	ANALYTE A	72.4	ng/mL	PERIOD 1	2013-01-19T19:30	PT12H
Bioinf-tese-01	PC	Bioinf-tese-01_01	9	A	ANALYTE A	24	ng/mL	PERIOD 1	2013-01-20T07:30	PT24H
Bioinf-tese-01	PC	Bioinf-tese-01_01	10	A	ANALYTE A	19.1	ng/mL	PERIOD 1	2013-01-20T19:22	PT36H
Bioinf-tese-01	PC	Bioinf-tese-01_01	11	B	ANALYTE B	BLQ	ng/mL	PERIOD 2	2013-01-26T07:15	-PT30M
Bioinf-tese-01	PC	Bioinf-tese-01_01	12	B	ANALYTE B	284.5	ng/mL	PERIOD 2	2013-01-26T08:30	PT1H
Bioinf-tese-01	PC	Bioinf-tese-01_01	13	B	ANALYTE B	315.9	ng/mL	PERIOD 2	2013-01-26T09:30	PT2H
Bioinf-tese-01	PC	Bioinf-tese-01_01	14	B	ANALYTE B	377.8	ng/mL	PERIOD 2	2013-01-26T10:30	PT3H
Bioinf-tese-01	PC	Bioinf-tese-01_01	15	B	ANALYTE B	335.4	ng/mL	PERIOD 2	2013-01-26T11:30	PT4H
Bioinf-tese-01	PC	Bioinf-tese-01_01	16	B	ANALYTE B	253	ng/mL	PERIOD 2	2013-01-26T13:30	PT6H
Bioinf-tese-01	PC	Bioinf-tese-01_01	17	B	ANALYTE B	150.3	ng/mL	PERIOD 2	2013-01-26T15:30	PT8H
Bioinf-tese-01	PC	Bioinf-tese-01_01	18	B	ANALYTE B	76.5	ng/mL	PERIOD 2	2013-01-26T19:30	PT12H
Bioinf-tese-01	PC	Bioinf-tese-01_01	19	B	ANALYTE B	41.7	ng/mL	PERIOD 2	2013-01-27T07:30	PT24H
Bioinf-tese-01	PC	Bioinf-tese-01_01	20	B	ANALYTE B	29.9	ng/mL	PERIOD 2	2013-01-27T19:30	PT36H

Figure 7 - Simplified Example of a PC dataset.

Regarding Figure 7, this figure represents a section of the PC dataset specific to the subject Bioinf-tese-01_01, where the values of the concentrations of both analytes A and B in the first five timepoints are shown. These values, along with the respective treatments, study periods and subjects, are later reported in a listing that will be included in the CSR, more specifically the section 16.2 which is the section specific for patient data listings (see chapter 4) [33].

3.2. ADaM Datasets

ADaM datasets are used to represent data used for analysis in a standardized format. In the process of developing these datasets, two distinct documents must be kept in mind, the ADaM Specification [37], and the ADaM Implementation Guide (ADaMIG) [38], which defines respectively the main guidelines and the implementation details that need to be strictly followed during the process. Just like in SDTM datasets, the software used to create ADaM datasets was the SAS Enterprise Guide 8.3.

Analysis Data Subject Level (ADSL) dataset

ADaM datasets can follow three kinds of structures. The first one is ADSL which is structured with one record per subject and includes variables such as demographic information, planned and

actual treatment variables, important dates, subject-level population flags, subgrouping variables, and other variables that describe a subject's experience in the clinical trial. The ADSL dataset must contain the required variables specified in the ADaM Implementation Guide (ADaMIG), as well as other subject-level variables that are important for analysis. Even if no other analysis datasets are submitted, ADSL and its related metadata are mandatory in a CDISC-based data submission from a clinical trial [37,38].

Regarding subject-level population flags, these variables indicate whether a subject belongs to a specific population or not. As examples we have the SAFFL, which is associated with the safety analysis population, the RANDFL, which indicates the population of randomized subjects, the COMPLFL, which identifies the subjects who completed the study, the PKPOPFL, which is associated with the pharmacokinetics analysis population, and the BEPOPFL, which is associated with the bioequivalence analysis population. The use of these flags varies from study to study, therefore documents such as the CSP and SAP must have the definition of the needed flags [37,38].

It is technically possible to include every single data value in a study as variables in a subject-level dataset such as ADSL. However, this is not the intended purpose of ADSL. Key endpoints and data that vary over time during the course of a study should be included in a BDS (Basic Data Structure) dataset instead [37,38].

STUDYID	USUBJID	SUBJID	SITEID	SEX	RACE	SAFFL	PKPOPFL	BEPOPFL	COMPLFL	RANDFL	ARM	TRTSEQA
Bioinf-tese-01	Bioinf-tese-01_01	1	BlueClinical Phase I	F	ASIAN	Y	Y	Y	Y	Y	Test-Reference	Test-Reference
Bioinf-tese-01	Bioinf-tese-01_02	2	BlueClinical Phase I	M	MULTIPLE	Y	Y	Y	Y	Y	Reference-Test	Reference-Test
Bioinf-tese-01	Bioinf-tese-01_03	3	BlueClinical Phase I	F	WHITE	Y	Y	N	N	Y	Test-Reference	Unplanned Treatment
Bioinf-tese-01	Bioinf-tese-01_04	4	BlueClinical Phase I	M	BLACK OR AFRICAN AMERICAN	Y	Y	Y	Y	Y	Reference-Test	Reference-Test
Bioinf-tese-01	Bioinf-tese-01_05	5	BlueClinical Phase I	F	WHITE	Y	Y	Y	Y	Y	Test-Reference	Test-Reference
Bioinf-tese-01	SCR_06	6	BlueClinical Phase I	M	WHITE	N	N	N	N	N	Not Assigned	Not Assigned
Bioinf-tese-01	SCR_07	7	BlueClinical Phase I	M	MULTIPLE	N	N	N	N	N	Screen Failure	Screen Failure

Figure 8 - Simplified Example of an ADSL dataset.

Basic Data Structure (BDS)

Speaking of BDS, this structure contains one or more records per subject, per analysis parameter, and per analysis timepoint, which is conditionally required depending on the analysis. In cases where there is no analysis timepoint, the structure is one or more records per subject per analysis parameter. The central set of variables in BDS describes the analysis parameter, including PARAM and related variables, and contains the analyzed value, such as AVAL and AVALC and related variables. Other variables in the dataset provide additional information about the analyzed value, such as subject identification, or describe and trace its derivation, such as DTYPE, or facilitate the analysis, such as treatment variables and covariates. When it comes to the variables' derivation, a rule must be followed which is described in the SAP document. Parametric and nonparametric

analyses are supported by the BDS. It is often recommended to have more than one BDS analysis dataset [37,38]. BDS also has a specific set of variables which are the --RFL. These flags determine if the subject was part of the specified analysis for a particular record. They are particularly helpful when the subject is included in the population at the subject level, but there are certain records in the dataset that do not meet the population requirements [37,38].

However, BDS does not support all statistical analyses. For example, it cannot perform simultaneous analysis of multiple dependent variables or correlation analysis across a range of response variables. BDS was not also designed to support analysis of incidence of adverse events or other occurrence data. The Pharmacokinetic Concentrations Analysis Dataset (ADPC) is an example of a dataset that follows the BDS structure [37,38].

STUDYID	USUBJID	SITEID	TRTA	EPOCH	AVISIT	ATPT	PARAM	PARAMCD	AVAL	PKPOPRFL	BEPOPRFL
Bioinf-tese-01	Bioinf-tese-01_01	BlueClinical Phase I	Drug A	PERIOD 1	Period 1: Confinement	PRE-DOSE	ANALYTE A (ng/mL)	A	0.00	Y	Y
Bioinf-tese-01	Bioinf-tese-01_01	BlueClinical Phase I	Drug A	PERIOD 1	Period 1: Confinement	1H POST-DOSE	ANALYTE A (ng/mL)	A	180.40	Y	Y
Bioinf-tese-01	Bioinf-tese-01_01	BlueClinical Phase I	Drug A	PERIOD 1	Period 1: Confinement	2H POST-DOSE	ANALYTE A (ng/mL)	A	436.60	Y	Y
Bioinf-tese-01	Bioinf-tese-01_01	BlueClinical Phase I	Drug A	PERIOD 1	Period 1: Confinement	3H POST-DOSE	ANALYTE A (ng/mL)	A	459.60	Y	Y
Bioinf-tese-01	Bioinf-tese-01_01	BlueClinical Phase I	Drug A	PERIOD 1	Period 1: Confinement	4H POST-DOSE	ANALYTE A (ng/mL)	A	428.50	Y	Y
Bioinf-tese-01	Bioinf-tese-01_01	BlueClinical Phase I	Drug A	PERIOD 1	Period 1: Confinement	6H POST-DOSE	ANALYTE A (ng/mL)	A	267.90	Y	Y
Bioinf-tese-01	Bioinf-tese-01_01	BlueClinical Phase I	Drug A	PERIOD 1	Period 1: Confinement	8H POST-DOSE	ANALYTE A (ng/mL)	A	181.20	Y	Y
Bioinf-tese-01	Bioinf-tese-01_01	BlueClinical Phase I	Drug A	PERIOD 1	Period 1: Confinement	12H POST-DOSE	ANALYTE A (ng/mL)	A	72.40	Y	Y
Bioinf-tese-01	Bioinf-tese-01_01	BlueClinical Phase I	Drug A	PERIOD 1	Period 1: Confinement	24H POST-DOSE	ANALYTE A (ng/mL)	A	24.00	Y	Y
Bioinf-tese-01	Bioinf-tese-01_01	BlueClinical Phase I	Drug A	PERIOD 1	Period 1: Ambulatory	36H POST-DOSE	ANALYTE A (ng/mL)	A	19.10	Y	Y
Bioinf-tese-01	Bioinf-tese-01_01	BlueClinical Phase I	Drug B	PERIOD 2	Period 2: Confinement	PRE-DOSE	ANALYTE B (ng/mL)	B	0.00	Y	Y
Bioinf-tese-01	Bioinf-tese-01_01	BlueClinical Phase I	Drug B	PERIOD 2	Period 2: Confinement	1H POST-DOSE	ANALYTE B (ng/mL)	B	284.50	Y	Y
Bioinf-tese-01	Bioinf-tese-01_01	BlueClinical Phase I	Drug B	PERIOD 2	Period 2: Confinement	2H POST-DOSE	ANALYTE B (ng/mL)	B	315.90	Y	Y
Bioinf-tese-01	Bioinf-tese-01_01	BlueClinical Phase I	Drug B	PERIOD 2	Period 2: Confinement	3H POST-DOSE	ANALYTE B (ng/mL)	B	377.90	Y	Y
Bioinf-tese-01	Bioinf-tese-01_01	BlueClinical Phase I	Drug B	PERIOD 2	Period 2: Confinement	4H POST-DOSE	ANALYTE B (ng/mL)	B	335.40	Y	Y
Bioinf-tese-01	Bioinf-tese-01_01	BlueClinical Phase I	Drug B	PERIOD 2	Period 2: Confinement	6H POST-DOSE	ANALYTE B (ng/mL)	B	253.00	Y	Y
Bioinf-tese-01	Bioinf-tese-01_01	BlueClinical Phase I	Drug B	PERIOD 2	Period 2: Confinement	8H POST-DOSE	ANALYTE B (ng/mL)	B	150.30	Y	Y
Bioinf-tese-01	Bioinf-tese-01_01	BlueClinical Phase I	Drug B	PERIOD 2	Period 2: Confinement	12H POST-DOSE	ANALYTE B (ng/mL)	B	76.50	Y	Y
Bioinf-tese-01	Bioinf-tese-01_01	BlueClinical Phase I	Drug B	PERIOD 2	Period 2: Confinement	24H POST-DOSE	ANALYTE B (ng/mL)	B	41.70	Y	Y
Bioinf-tese-01	Bioinf-tese-01_01	BlueClinical Phase I	Drug B	PERIOD 2	Period 2: Ambulatory	36H POST-DOSE	ANALYTE B (ng/mL)	B	29.90	Y	Y

Figure 9 - Simplified Example of an ADPC dataset.

Occurrence Data Structure (OCCDS)

At last, we have the OCCDS datasets, which are done for the purpose of analyzing occurrence data, such as adverse events.

To utilize this structure, certain criteria must be satisfied. Firstly, AVAL and AVALC are unnecessary for the analysis. Instead, occurrences are tallied, and multiple records for each assessment are typically present [39].

Secondly, a dictionary is employed to encode these occurrences, and it should not be remapped to BDS variables PARAM and generic -*CAT variables. Such remapping would result in the loss of the dictionary's structure and meaning. Adverse events and concomitant medications usually require a dictionary, and it is recommended for medical history as well. Although less common, clinical events, procedures, and substance use may also be coded using a dictionary [39].

Lastly, the data content is typically unaltered for analysis purposes. Therefore, there is no need for analysis versions of variables that store the dictionary hierarchy or category terms [39].

STUDYID	USUBJID	AESQ	AETERM	AESTDTC	AEENDTC	EPOCH	TRTEMFL	PREFL	SAFFL	AOCFL	AOCCFL	AOCCSFL	AOCCPFL	AESEV	AEREL
Bioinf-tese-01	Bioinf-tese-01_01	1	Somnolence	2013-01-19T07:50	2013-01-19T15:30	PERIOD 1	Y		Y	Y	Y	Y	Y	MODERATE	REASONABLY POSSIBLE
Bioinf-tese-01	Bioinf-tese-01_01	2	Nausea	2013-01-19T10:00	2013-01-19T15:00	PERIOD 1	Y		Y		Y	Y	Y	MILD	REASONABLY POSSIBLE
Bioinf-tese-01	Bioinf-tese-01_02	1	Somnolence	2013-01-19T09:00	2013-01-19T15:30	PERIOD 1	Y		Y	Y	Y	Y	Y	MODERATE	REASONABLY POSSIBLE
Bioinf-tese-01	Bioinf-tese-01_02	2	Soft stools	2013-01-19T11:00	2013-01-19T11:00	PERIOD 1	Y		Y		Y	Y	Y	MILD	REASONABLY POSSIBLE
Bioinf-tese-01	Bioinf-tese-01_03	1	Somnolence	2013-01-19T08:30	2013-01-19T16:00	PERIOD 1	Y		Y	Y	Y	Y	Y	MILD	REASONABLY POSSIBLE
Bioinf-tese-01	Bioinf-tese-01_03	2	Flatulence	2013-01-19T11:00	2013-01-19T20:00	PERIOD 1	Y		Y		Y	Y	Y	MILD	REASONABLY POSSIBLE
Bioinf-tese-01	Bioinf-tese-01_03	3	Flu like syndrome	2013-01-22T16:00	2013-01-23T07:00	WASHOUT	Y		Y		Y	Y	Y	MILD	REASONABLY POSSIBLE
Bioinf-tese-01	Bioinf-tese-01_04	1	Rhinopharyngitis	2013-01-03T06:00	2013-01-03T18:00	SCREENING		Y	Y					MILD	NOT REASONABLY POSSIBLE
Bioinf-tese-01	Bioinf-tese-01_04	2	Somnolence	2013-01-19T08:30	2013-01-19T16:00	PERIOD 1	Y		Y	Y	Y	Y	Y	MILD	REASONABLY POSSIBLE
Bioinf-tese-01	Bioinf-tese-01_04	3	Constipation	2013-01-26T15:00	2013-01-26T06:00	PERIOD 2	Y		Y		Y	Y	Y	MILD	REASONABLY POSSIBLE
Bioinf-tese-01	Bioinf-tese-01_05	1	Dry cough	2013-01-16T09:00	2013-01-19T12:00	SCREENING		Y	Y					MILD	NOT REASONABLY POSSIBLE
Bioinf-tese-01	Bioinf-tese-01_05	2	Somnolence	2013-01-19T08:30	2013-01-19T18:00	PERIOD 1	Y		Y	Y	Y	Y	Y	MILD	REASONABLY POSSIBLE
Bioinf-tese-01	Bioinf-tese-01_05	3	Frontal headache	2013-01-19T16:00	2013-01-20T06:00	PERIOD 1	Y		Y			Y	Y	MILD	REASONABLY POSSIBLE
Bioinf-tese-01	Bioinf-tese-01_05	4	Somnolence	2013-01-26T08:45	2013-01-26T19:00	PERIOD 2	Y		Y					MODERATE	REASONABLY POSSIBLE

Figure 10 - Simplified Example of an ADAE dataset.

Regarding the Adverse Events Analysis Dataset (ADAE) represented in Figure 12, it is important to make sure that all the AEs that will be used for analysis are represented. As such, flags such as the TRTEMFL, which indicates treatment emergent AEs (TEAEs), and PREFL, which indicates the AEs that occurred prior to the exposure to the treatment, are used to accomplish the AEs representation. These flags must also be considered when making TFLs focused on AEs statistics [39].

4. TFLs

This chapter will focus on the main topic of this report, which are the TFLs, where firstly each type of TFLs is described regarding the type of data represented and how the data is reported, and lastly, examples of the process of creating a table and a figure will be demonstrated, which are done according with the regulatory requirements mentioned in the chapter 2.2. The software used to create the mentioned TFLs was the SAS Enterprise Guide 8.3, which runs on SAS 9.4.

TFLs, or Tables, Figures and Listings, are important components of a Clinical Study Report (CSR), as they represent the results and analysis of a clinical trial in a standardized format for regulatory submissions and publications, in order to guarantee the transparency of the results.

Tables are used to present numerical data in an organized and concise manner. They are typically used to summarize study results and present statistical analyses. Tables can include information such as demographic characteristics of study participants, baseline characteristics, efficacy and safety outcomes, and adverse events [40]. The process of making a table will be demonstrated in the chapter 4.1, as well as an example resulted from the process.

In contrast, figures provide visual representations of clinical data and typically take the shape of graphs, charts, or diagrams. Figures can be used to present data that is difficult to express in a table format, such as changes in plasma concentration of a treatment over time [40]. The process of making a figure will be demonstrated in the chapter 4.2, as well as an example resulted from the process.

Although listings and tables are fairly similar, tables typically give a summary of the data or a statistical analysis, whereas listings typically present a more complete description of the individual data points or observations that were gathered throughout the study. Listings and can be used by regulators or other researchers to verify the accuracy of the data presented in the tables and figures. Listings can be very large and include hundreds or even thousands of individual data points [40]. The process of creating a listing won't be shown, since it is very similar to the process of creating a table.

Listing 1 - Individual Pharmacokinetic Concentration Data.

Subject No.	Period	Investigational Product	Pre-dose	Time Post-Dose (h)								
				1.00	2.00	3.00	4.00	6.00	8.00	12.00	24.00	36.00
Bioinf-tese-01_01	Period 1	Drug A	BLQ	180.4	436.6	459.6	428.5	267.9	181.2	72.4	24.0	19.1
	Period 2	Drug B	BLQ	284.5	315.9	377.8	335.4	253.0	150.3	76.5	41.7	29.9
Bioinf-tese-01_02	Period 1	Drug B	BLQ	113.2	550.9	609.4	557.5	448.5	297.5	162.6	63.9	36.9
	Period 2	Drug A	BLQ	122.5	709.5	671.9	610.1	482.4	308.8	167.3	68.5	19.4
Bioinf-tese-01_03	Period 1	Drug A	BLQ	337.7	451.6	493.3	467.6	359.2	252.3	132.6	41.7	33.8
Bioinf-tese-01_04	Period 1	Drug B	BLQ	327.9	384.4	393.6	385.0	331.7	253.9	186.5	111.2	86.6
	Period 2	Drug A	BLQ	74.3	156.1	225.6	416.7	633.9	444.8	234.9	99.2	56.2
Bioinf-tese-01_05	Period 1	Drug A	BLQ	67.4	413.2	443.2	417.3	307.5	226.5	103.0	35.4	20.2
	Period 2	Drug B	BLQ	428.0	457.0	486.5	457.2	299.1	179.7	74.0	26.3	18.9

Listing 2 - Individual Adverse Events Concentration Data.

Subject No.	Reported Term	Serious?	Maximal Severity	Causality	Action Taken	Medication Required?	Outcome	Adverse Event Start Date and Time	Adverse Event End Date and Time
Bioinf-tese-01_01	Somnolence	No	Moderate	Reasonably Possible	Dose Not Changed	No	Recovered/ Resolved	19JAN2013 07:50	19JAN2013 15:30
	Nausea	No	Mild	Reasonably Possible	Dose Not Changed	No	Recovered/ Resolved	19JAN2013 10:00	19JAN2013 15:00
Bioinf-tese-01_02	Somnolence	No	Moderate	Reasonably Possible	Dose Not Changed	No	Recovered/ Resolved	19JAN2013 09:00	19JAN2013 15:30
	Soft stools	No	Mild	Reasonably Possible	Not Applicable	No	Recovered/ Resolved	19JAN2013 11:00	19JAN2013 11:00
Bioinf-tese-01_03	Flu like syndrome	No	Mild	Reasonably Possible	Drug Withdrawn	No	Recovered/ Resolved	22JAN2013 16:00	23JAN2013 07:00
	Flatulence	No	Mild	Reasonably Possible	Dose Not Changed	No	Recovered/ Resolved	19JAN2013 11:00	19JAN2013 20:00
Bioinf-tese-01_04	Somnolence	No	Mild	Reasonably Possible	Dose Not Changed	No	Recovered/ Resolved	19JAN2013 08:30	19JAN2013 16:00
	Constipation	No	Mild	Reasonably Possible	Dose Not Changed	No	Recovered/ Resolved	26JAN2013 15:00	26JAN2013 06:00
Bioinf-tese-01_05	Rhinopharyngitis	No	Mild	Not Reasonably Possible	Not Applicable	Yes	Recovered/ Resolved	03JAN2013 06:00	03JAN2013 18:00
	Frontal headache	No	Mild	Reasonably Possible	Dose Not Changed	No	Recovered/ Resolved	19JAN2013 16:00	20JAN2013 06:00
Bioinf-tese-01_05	Somnolence	No	Mild	Reasonably Possible	Dose Not Changed	No	Recovered/ Resolved	19JAN2013 08:30	19JAN2013 18:00
	Somnolence	No	Moderate	Reasonably Possible	Not Applicable	No	Recovered/ Resolved	26JAN2013 08:45	26JAN2013 19:00
	Dry cough	No	Mild	Not Reasonably Possible	Not Applicable	No	Recovered/ Resolved	16JAN2013 09:00	19JAN2013 12:00

4.1. Creation of a Table

In this section, the process of producing a Table will be demonstrated and explained. In this case, the table produced will show a summary of the treatment-emergent adverse events (TEAEs), which are the AEs that occurred after the dosage of treatment has started, will be created.

For this specific table, the first step is to associate the TEAEs with the respective treatment (in this case, either test or reference). This is done by merging the ADAE dataset with the EX and SE datasets from SDTM with the aid of the SAS proc SQL statement, just as shown below:

```
proc sql;
  create table work._TEMP1 as
  select distinct
    case
  when ex2.exstdtc~='' AND ae.aestdrc >= ex2.exstdtc then 'PERIOD 2'
  when ex1.exstdtc~='' AND ae.aestdrc >= ex1.exstdtc then 'PERIOD 1'
  else 'SCREENING' end as trtepoeh,
  ae.*
  from adam.adae as ae
  left join sdtm.ex as ex1 on (ae.usubjid=ex1.usubjid and ex1.epoch = 'PERIOD
1')
  left join sdtm.ex as ex2 on (ae.usubjid=ex2.usubjid and ex2.epoch = 'PERIOD
2')
  where ae.TRTEMFL='Y';

  create table work._AETRT (drop=trtepoeh0) as
  select se.etcd as trt,
    ae.trtepoeh0 as trtepoeh,
    ae.*
  from work._TEMP1 (rename=(trtepoeh=trtepoeh0)) as ae
  left join sdtm.se as se on (se.usubjid=ae.usubjid and se.epoch
=upcase(ae.trtepoeh0))
  where se.etcd in ('A', 'B')
  UNION ALL
  select se2.etcd as trt,
    se2.epoch as trtepoeh,
    ae.*
  from work._TEMP1 as ae
  left join sdtm.se as se on (se.usubjid=ae.usubjid and
se.epoch=upcase(ae.trtepoeh))
  left join sdtm.se as se2 on (se2.usubjid=ae.usubjid)
  where se2.taetord<se.taetord and se.epoch='WASHOUT'
  group by ae.usubjid, ae.aeseq
  having se2.taetord=max(se2.taetord);
run;
```

Code 1 - Association of TEAEs with the study treatments.

The next step serves to determine the number of subjects that took each of the treatments. Just like in the first step, the proc SQL statement will be used:

```
proc sql;
  create table work._formSubj as
  select se.trt as formul,
    count(distinct ex.usubjid) as expformul
  from tsdtm.ex as ex
```

```

        inner join (select usubjid, epoch, etcd as trt from tsdtm.se) as se on
(se.usubjid=ex.usubjid and se.epoch=ex.epoch)
        group by se.trt;
run;

```

Code 2 - Number of Subjects in each treatment.

The following step starts the summarization of the TEAEs proper, where it's taken to account the N° of subjects, the N° of TEAEs, the MedDRA definitions for the System Organ Class (SOC) and the Preferred Term (PT) and the relation. This step can be divided into four parts. The first part will summarize the TEAEs for each treatment, the second will summarize the TEAEs in total, the third part will summarize the TEAEs regarding their severity, while the fourth part unites the datasets produced in the parts mentioned before. For each of the parts, a proc SQL statement will be used, apart from the fourth part, where a data statement and a proc sort statement are used.

```

proc sql;
create table work._test1 as
SELECT 'Number of Subjects {Nr. of TEAEs} (% of Subjects)' length=200 as soc,
      0 as socorder,
      '' as pt,
      0 as ptorder,
      catt(put(fs.formul,$_treat.), ' (n=', fs.expformul,')') as formul,
      fs.expformul,
      (case when fs.formul = 'B' then 1
            when fs.formul = 'A' then 0 else -1 end) as headorder,
      'EMERGENT' as type,
      count(distinct ae.usubjid) as subjno,
      count(ae.aeseq) as aeno,
      '1st query' length=20 as query
      FROM work._formSubj as fs
      INNER JOIN work._AETRT as ae on (ae.trt=fs.formul)
      GROUP BY fs.formul, fs.expformul
UNION
SELECT 'Number of Subjects {Nr. of TEAEs} (% of Subjects)' length=200 as soc,
      0 as socorder,
      '' as pt,
      0 as ptorder,
      catt(put(fs.formul,$_treat.), ' (n=', fs.expformul,')') as formul,
      fs.expformul,
      (case when fs.formul = 'B' then 1
            when fs.formul = 'A' then 0 else -1 end) as headorder,
      'RELATED' as type,
      count(distinct ae.usubjid) as subjno,
      count(ae.aeseq) as aeno,
      '2nd query' as query
      FROM work._formSubj as fs
      LEFT JOIN work._AETRT as ae on (ae.trt=fs.formul and ae.aerel in
('REASONABLY POSSIBLE'))
      GROUP BY fs.formul, fs.expformul;
run;

```

Code 3 - Portion of query that summarizes the TEAEs regarding each study treatment.

```

proc sql;
create table work._test2 as
SELECT DISTINCT 'Number of Subjects {Nr. of TEAEs} (% of Subjects)'
length=200 as soc,
      0 as socorder,
      '' as pt,

```

```

        0 as ptorder,
        totals.formul,
        totals.expformul,
        2 as headorder,
        'EMERGENT' as type,
        count(distinct ae.usubjid) as subjno,
        count(*) as aeno
    FROM work._formSubj as fs
    LEFT JOIN work._totals as totals on (1 = 1)
    INNER JOIN work._AETRT as ae on (ae.trt=fs.formul) UNION ALL
    SELECT DISTINCT 'Number of Subjects {Nr. of TEAEs} (% of Subjects)' length=200
as soc,
        0 as socorder,
        '' as pt,
        0 as ptorder,
        totals.formul,
        totals.expformul,
        2 as headorder,
        'RELATED' as type,
        count(distinct ae.usubjid) as subjno,
        count(ae.aeseq) as aeno
    FROM work._formSubj as fs
    LEFT JOIN work._totals as totals on (1 = 1)
    LEFT JOIN work._AETRT as ae on (ae.trt=fs.formul and ae.aerel in
('REASONABLY POSSIBLE'))
run;

```

Code 4 - Portion of query that summarizes the TEAEs in total.

```

proc sql;
    create table work._test3 as
    SELECT DISTINCT 'Severity' length=200 as soc,
    2 as socorder,
    CASE WHEN ae.aesev="NA" THEN ae.aesev ELSE propcase(ae.aesev) END as pt,
    1 as ptorder,
    totals.formul,
    totals.expformul,
    2 as headorder,
    'EMERGENT' as type,
    count(distinct ae.usubjid) as subjno,
    count(*) as aeno,
    '1st sev query' length=20 as query
    FROM work._formSubj as fs
    LEFT JOIN work._totals as totals on (1 = 1)
    INNER JOIN work._AETRT as ae on (ae.trt=fs.formul)
    GROUP BY ae.aesev
    UNION ALL
    SELECT DISTINCT 'Severity' length=200 as soc,
    2 as socorder,
    CASE WHEN ae.aesev="NA" THEN ae.aesev ELSE propcase(ae.aesev) END as pt,
    1 as ptorder,
    totals.formul,
    totals.expformul,
    2 as headorder,
    'RELATED' as type,
    count(distinct ae.usubjid) as subjno,
    count(*) as aeno,
    '2nd sev query' as query
    FROM work._formSubj as fs
    INNER JOIN work._totals as totals on (1 = 1)
    INNER JOIN work._AETRT as ae on (ae.trt=fs.formul and ae.aerel in
('REASONABLY POSSIBLE'))
    GROUP BY ae.aesev
run;

```

Code 5 - Portion of query that summarizes the TEAEs regarding their severity.

```

data work._test;
    set _test1 _test2 _test3;
run;
proc sort data=work._test;
    by headorder;
run;

```

Code 6 - Merging of each summarization dataset into one.

For the final step, the goal is to export the summarized data into a document file. As such, statements such as proc template, proc report and Output Delivery System (ODS) can be used for this procedure.

Regarding proc template, this statement is used as a tool to customize the style format of a TFL. Details of a document, or a specific part of a document, such as the font style, font size, text alignment, background color, borders and margins size can be specified in the proc template statement, just as shown in the code below:

```

proc template;
    define style TFLSTYLE;
        style data /
            fontfamily="Times New Roman"
            fontsize=12pt;
        style header /
            fontfamily="Times New Roman"
            fontsize=12pt;
        style NoteContent /
            fontfamily="Times New Roman"
            fontsize=10pt;
        style PageNo from PageNo/
            font_size=0.1pt
            background=white
            foreground=white;
        style SystemFooter from SystemFooter /
            fontfamily="Times New Roman"
            fontsize=10pt
        style systemtitle /
            fontfamily="Times New Roman"
            fontsize=11pt
            fontweight=bold;
        style table /
            fontfamily="Times New Roman"
            borderwidth=0
            outputwidth=100%
            height=60%
            rules=rows;
        style GraphFonts /
            'GraphDataFont' = ("Times New Roman", 7pt)
            'GraphLabelFont' = ("Times New Roman", 10pt)
            'GraphFootnoteFont' = ("Times New Roman", 10pt);
        style body from document /
            leftmargin=1.25in
            rightmargin=0.78in
    end;
run;

```

Code 7 - Simplified query for the creation of the TFLSTYLE custom format style.

When it comes to the proc report, this statement is used for creating tabular reports from data sets. It is a flexible and powerful reporting tool that allows full customization of a report. Some of the features of proc report include output control, where the variables to be included or excluded are specified, ability to format the output, where just like in proc template, details such as font size, text alignment, borders and margins can be customized, grouping and summarization of variables, ordering and sorting of variables, ability to specify the titles, footnotes and headings and computing of new variables to report.

In order to export the report made by the proc report statement to a document file, the ODS statement must be used. This SAS feature enables the creation and control of the output of SAS procedures and programs, directing it to the desired destination, like a file for example. Just like proc template and proc report, ODS uses a set of options to control the output format and appearance, including titles, footnotes, colors, fonts, and other formatting elements. In this particular case, the report created by proc report will be exported to a Rich Text Format (RTF) file and will follow the format style defined by the proc template mentioned before.

```
options orientation=landscape;
ods rtf file="(file path)\Table.rtf" style=TFLSTYLE;
title j=left font= 'Times New Roman' bold height=11pt 'Summary of Treatment-
Emergent Adverse Events (TEAEs)';

proc report data=work._test(where=(soc ~= '')) headskip headline split="S"
style(report)=[just=l fontsize=10pt] style(column)=[fontsize=10pt ]
style(lines)=[just=left fontweight=bold fontsize=10pt ]
style(header)=[fontsize=10pt fontweight=bold just=center ];
    columns (({fS={bordertopcolor=black borderbottomcolor=white} ' '
'fS={bordertopcolor=white borderbottomcolor=black} ' (socorder soc pt ptorder
dispsocpt))(formul, type, (expformul subjno aeno value))) ;

    define socorder / group ' ' style(header)=[width=20px just=left] noprint;
    define soc / group missing ' ' style(header)=[width=20px just=left asis=on
bordertopwidth=0px] noprint;
    define pt / group missing ' ' style(header)=[width=20px just=left] noprint;
    define ptorder / group ' ' noprint;
    define formul / across ' ' order=data;
    define type / across ' ';
    define expformul / analysis ' ' sum noprint;
    define subjno / analysis ' ' sum noprint;
    define aeno / analysis ' ' sum noprint;
    define value / computed ' ' format=$w.;
    define dispsocpt / computed ' ' style(header)=[bordertopcolor=white
bordertopwidth=1px] format=$w.;
    compute dispsocpt / char length=100;
        if socorder~='' and socorder~='last_socorder then do;
            last_socorder = lag(socorder);
        end;
        if socorder=2 or last_socorder=2 then do;
            call define(_row_, "style", "STYLE=[fontweight=bold]");
            dispsocpt=cats(" ",pt);
        end;
    else if ptorder=0 then do;
        call define(_row_, "style", "STYLE=[fontweight=bold]");
```

```

        dispsocpt=soc;
    end;
    else do;
        dispsocpt=cats("    ",pt);
    end;
endcomp;

break after soc / skip;

compute before soc;
    count=0;
    length text $60;
    text = '';
    num = 0;
    space = '';
    num2 = 0;
    num3 = 0;
    if socorder = 2 then do;
        text = '£S={bordertopcolor=black bordertopwidth = 1px}
Severity';
        space = '  ';
        num = 60;
        num3 = 0;
        num2 = 1;
    end;
    line space $VARYING. num2;
    line space $VARYING. num3;
    line text $VARYING. num;
    line space $VARYING. num3;
endcomp;

compute after soc;
    length text $60;
    if socorder = 0 then do;
        text = 'MedDRA';
        text2 = 'System Organ Class (SOC)';
        text3 = '£S={font_weight=medium} Preferred Term (PT)';
        text4 = '';
        num = 50;
        num3 = num;
        num2 = 1;
    end;
    line text4 $VARYING. num2;
    line text $VARYING. num;
    line text2 $VARYING. num;
    line text3 $VARYING. num3;

run;

ods rtf close;

```

Code 8 - Portion of the query used for the report of the summarized TEAEs and exportation to a RTF file.

After running the code, the following table is produced, which just like was said before, summarizes the TEAEs that occurred during the study. More specifically, this table shows information about the nº of subjects, nº of TEAEs and the percentage of the subjects that suffered a TEAE, while regarding factors such as the treatments the subjects received, the relation of the TEAEs with the treatments, the MedDRA coding respective of each TEAE and the TEAEs' severity.

Table 1 - Summary of TEAEs.

	Test (n=5)		Reference (n=4)		Total (N=5)	
	All TEAEs	Drug-related TEAEs	All TEAEs	Drug-related TEAEs	All TEAEs	Drug-related TEAEs
Number of Subjects {Nr. of TEAEs} (% of Subjects)	4 {8} (80%)	4 {8} (80%)	3 {4} (75%)	3 {4} (75%)	5 {12} (100%)	5 {12} (100%)
MedDRA						
System Organ Class (SOC)						
Preferred Term (PT)						
Gastrointestinal disorders	3 {3} (60%)	3 {3} (60%)	1 {1} (25%)	1 {1} (25%)	4 {4} (80%)	4 {4} (80%)
Constipation	1 {1} (20%)	1 {1} (20%)	0	0	1 {1} (20%)	1 {1} (20%)
Faeces soft	0	0	1 {1} (25%)	1 {1} (25%)	1 {1} (20%)	1 {1} (20%)
Flatulence	1 {1} (20%)	1 {1} (20%)	0	0	1 {1} (20%)	1 {1} (20%)
Nausea	1 {1} (20%)	1 {1} (20%)	0	0	1 {1} (20%)	1 {1} (20%)
General disorders and administration site conditions	1 {1} (20%)	1 {1} (20%)	0	0	1 {1} (20%)	1 {1} (20%)
Influenza like illness	1 {1} (20%)	1 {1} (20%)	0	0	1 {1} (20%)	1 {1} (20%)
Nervous system disorders	3 {4} (60%)	3 {4} (60%)	3 {3} (75%)	3 {3} (75%)	5 {7} (100%)	5 {7} (100%)
Headache	1 {1} (20%)	1 {1} (20%)	0	0	1 {1} (20%)	1 {1} (20%)
Somnolence	3 {3} (60%)	3 {3} (60%)	3 {3} (75%)	3 {3} (75%)	5 {6} (100%)	5 {6} (100%)
Severity						
Mild	4 {7} (80%)	4 {7} (80%)	2 {2} (50%)	2 {2} (50%)	5 {9} (100%)	5 {9} (100%)
Moderate	1 {1} (20%)	1 {1} (20%)	2 {2} (50%)	2 {2} (50%)	3 {3} (60%)	3 {3} (60%)

4.2. Creation of a Figure

In this chapter, the process of making a figure will be shown and interpreted. To be precise, this method will create a line chart that not only compares the arithmetic mean plasma concentrations between Drug A and Drug B for each study timepoint, but also shows the evolution of the arithmetic mean plasma concentrations of each drug during the course of the study. It should be mentioned that this chart will follow a linear scale.

The process of creating the chart can be easily divided into two steps. The first step involves the calculation of arithmetic mean plasma concentrations for both Drug A and Drug B., where the main statements would be `proc sql`, in order to fetch the required data for the calculations, and `proc summary`, which is a statement that calculates summary statistics, such as the mean. In this case, `proc summary` is used to calculate the means of the plasma concentrations, taking into account the investigational product and the study timepoints.

```
proc sql;
    create table work._temp as
    select pc.usubjid,
           pc.pctested as drug,
           pc.pctest as trt,
           pc.pcptnum format=best.,
           case when pc.pcstresc='BLQ' then 0 else pc.pcstresn end as linear
    from tsdtm.pc as pc
    where pc.usubjid in (select usubjid from tadam.adpp where BEPOPRFL="Y")
    order by drug, trt, pctptnum;
run;

proc summary data=work._temp noprint;
    var linear;
    by drug trt pctptnum;
    output out=work.PCSTATS(drop=_type_ _freq_) mean=;
run;
```

```

proc sort data=work.PCSTATS;
    by drug pctptnum;
run;

proc transpose data=work.PCSTATS out=work.PCSTATS1 (rename=(ANALYTE_B=B
ANALYTE_A=A) drop=_name_);
    by drug pctptnum;
    id trt;
    var linear;
run;

```

Code 9 - Calculation of arithmetic mean plasma concentrations.

The last step involves the creation of the chart itself, where the main statement is the proc sgplot. This statement can create one or multiple plots at the same time and overlap them in the same graph. It can create simple plots such as a line chart as well as a statistical graph such as a regression plot. Proc sgplot also has defined options that allow the customization of a chart, making it possible to modify its appearance and had additional features to it such as legends. Just like the table before, the chart produced will be exported to a rtf file with the ods statement.

```

ods rtf file="(file path)\Chart.rtf" style = TFLSTYLE bodytitle nogtitle;
title 'Arithmetic Mean Plasma Concentration Versus Time Profiles Following
Administration of Analytes A and B.';

PROC SGPLOT DATA = WORK._ PCSTATS1 pad = (bottom = 2pct left=10pct) NOAUTOLEGEND
NOBORDER ;

    xaxis label='Time (h)' values=(0 to 36 by 4) LABELATTRS=(Color=Black
Family="Times New Roman" Size=12 Weight=Bold) valueattrs = (size = 12pt
Family="Times New Roman" weight=bold);
    yaxis label='Analytes Concentration (ng/ml)' LABELATTRS=(Color=Black
Family="Times New Roman" Size=12 Weight=Bold) valueattrs = (size = 12pt
Family="Times New Roman" weight=bold) min=0 minorcount=3 THRESHOLDMAX=1;
    series x=PCTPTNUM y=A / markers lineattrs = (color = blue thickness = 0.5
pattern = solid) markerattrs=(symbol=CircleFilled color=blue size=7) name="A"
legendlabel="ANALYTE A";
    series x=PCTPTNUM y=B / markers lineattrs = (color = black thickness = 0.5
pattern = solid) markerattrs=(symbol=Circle color=black size=7) name="B"
legendlabel="ANALYTE B";
    keylegend "B" "A" / valueattrs=(color=black Family="Times New Roman" size=12pt
weight=bold) noborder OUTERPAD=(top=50px);

RUN;
ods rtf close;

```

Code 10 - Simplified Query used for Creation of the chart that presents the arithmetic mean plasma concentrations and exportation to RTF file.

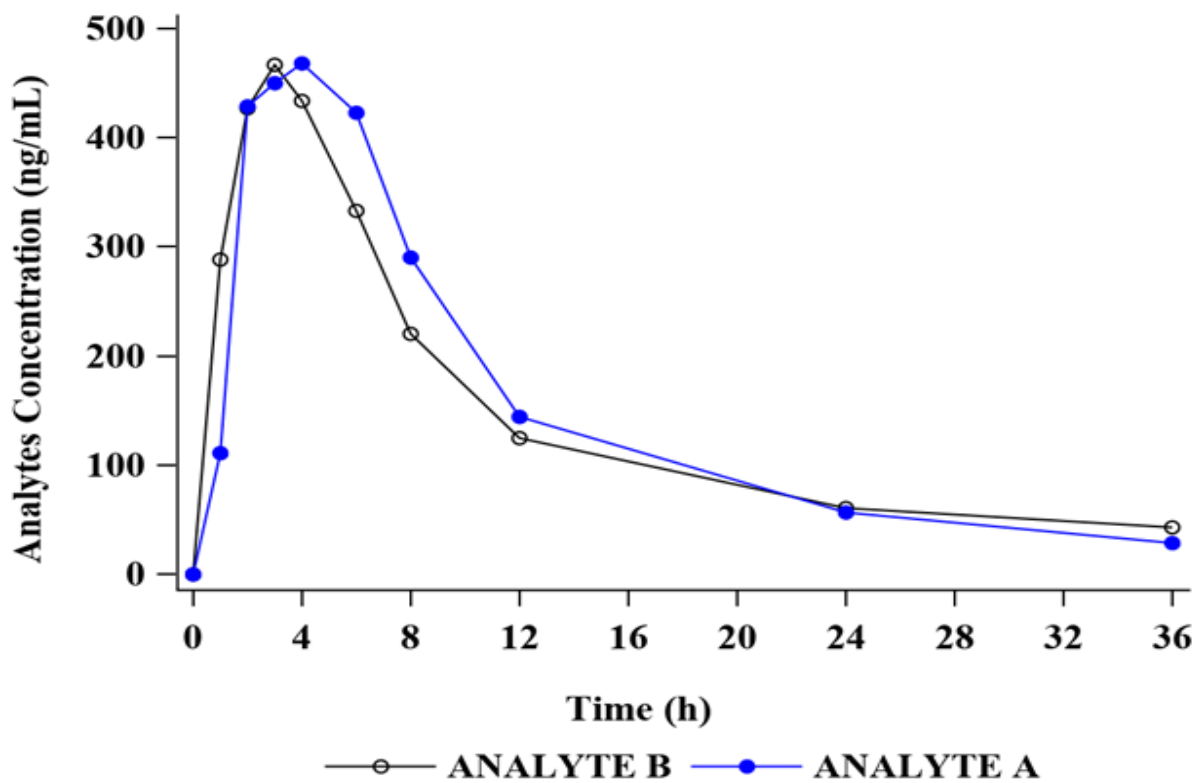


Figure 11 - Arithmetic Mean Plasma Concentration Versus Time Profiles Following Administration of Analytes A and B.

5. Discussion

As seen before, the production of TFLs is a crucial step for the submission of a study since they're a major component of the CSR and are needed to represent the results and data analysis performed during the study. However, the general process of making TFLs has some issues that need to be discussed. Firstly, each study exhibits certain variations, leading to discrepancies between what the team of medical writing desire and what is actually produced for the CSR. Even a slight difference requires additional time for rewriting and validating the program, often resulting in the creation and validation of entirely new programs [41].

Moreover, any minor changes in the output necessitate updates to the programs, subsequent revalidation, and the need to recheck the output for consistency in numerical values across all tables. For instance, a fundamental metric like the number of patients in the treated set might need to be calculated numerous times throughout a study. Therefore, it is crucial to ensure the accuracy and consistency of these numbers across all tables through meticulous verification [41].

One possible solution to these problems is the creation of an application that, with the usage of metadata, fills automatically TFL shells with specific results. Applications for TFLs' creation are

already widely accepted in numerous industries. While the pharmaceutical sector has been relatively slow to adopt this approach, there is no inherent barrier preventing its implementation, as long as a clear traceability can be established from raw data to the analysis results presented in any subsequent TFL. This traceability ensures that the need to reprogram an analysis, which has already been programmed elsewhere within the organization, can be eliminated. Thanks to the automated generation through the use of metadata, a significant reduction in the time needed to generate standard tables can be achieved. This time-saving benefit allows for a more extensive review of the data, ensuring its cleanliness, while also creating additional opportunities to explore new and innovative analyses [41].

During the course of the internship, I received training regarding most of the procedures of CDP, in order for me to get familiarized with them as well as the regulatory requirements involved. During this period, I also received training regarding other departments such as the Data Management (DM) department, where I was taught the processes conducted by DM, the Pharmacometrics (PK) department, where I received training on the principles of pharmacokinetics and on the statistical analysis performed in Bioavailability/Bioequivalence studies, and many others with which CDP normally interacts with during the procedures. It was only during the beginning of the year when I finally started doing CDP tasks that were deemed official, and at the beginning of May, I ended the training of CDP, being officially deemed adapted to perform any CDP procedure. It must be said that some of the mandatory training was done outside of BlueClinical's scope, such as Wizer's online course in Security Awareness, SYNEOS Health's online course in GCPs, and the SAS online training courses which introduced me to the SAS programming language. More information regarding the training performed during the internship can be found in table 2.

Table 2 - Training performed during internship.

Completion Date (dd-mmm-yyyy)	Total Duration (hh:mm)	Training	
		Title	Training Institution
04-OCT-2022	00:30	Pharmacovigilance welcome training	BlueClinical
04-OCT-2022	01:00	Welcome Training Plan: Quality Assurance & Quality Management System	BlueClinical
04-OCT-2022	01:00	Welcome training: Human Resources	BlueClinical
04-OCT-2022	01:00	Welcome training: Introduction to the company	BlueClinical
04-OCT-2022	01:50	Welcome training: Information & Technologies	BlueClinical
18-OCT-2022	21:00	SAS Programming 1: Essentials	SAS

19-OCT-2022	01:00	Security Awareness	Wizer
19-OCT-2022	02:00	Good Clinical Practice: A Refresher Course for all Site Personnel Working on Clinical Research Studies	SYNEOS Health
24-OCT-2022	21:00	Statistics 1: Introduction to ANOVA, Regression, and Logistic Regression	SAS
26-OCT-2022	00:45	Initial training in the General Data Protection Regulation	BlueClinical
07-NOV-2022	00:30	Ergonomia – Prevenção de Riscos Profissionais	BlueClinical
11-NOV-2022	01:00	Introduction to Pharmacokinetics: Drug Absorption and Bioequivalence Principles	BlueClinical
11-NOV-2022	01:30	Statistical Analysis of Bioavailability/Bioequivalence (BA/BE) Studies	BlueClinical
14-NOV-2022	01:00	Clinical Study Process	BlueClinical
24-NOV-2022	02:30	Clinical Data Management Processes	BlueClinical
30-NOV-2022	00:05	Handover – Transferência de Responsabilidades	BlueClinical
05-DEC-2022	01:00	Training Binder – Preenchimento e Manutenção	BlueClinical
26-JAN-2023	00:30	Electronic Signature Terms of Use	BlueClinical
01-FEB-2023	14:00	SAS Macro Language 1: Essentials	SAS
22-FEB-2023	01:00	Plano de Continuidade de Negócios	BlueClinical
28-FEB-2023	00:58	Como funciona a hemodiálise	BlueClinical
04-MAY-2023	01:00	Bioanalytical Laboratory Communications	BlueClinical
05-MAY-2023	38:00	Clinical Data Programming Activities	BlueClinical

Regarding my opinion on the internship, it allowed me to refresh some of the knowledge that was gained during the first year in the Master of Clinical Bioinformatics, such as basic concepts of general programming and clinical research, but it also made me gain more knowledge regarding specific processes in clinical research and SAS, which was a completely new programming language for me. Not only that, but this internship forced me to develop some Soft Skills like time management, priority setting and organization that are highly recommended to perform the various CDP procedures. However, one of my biggest expectations of the internship was that I could also develop some skills regarding statistical analysis and data visualization in clinical research, which due to the nature of the CDP processes, that didn't really happen, which I find somewhat disappointing, since I think that a CDP should also have a good background in statistics. Therefore, I think that in future internships to be conducted in BlueClinical's CDP department, the training should include activities more focused on the field of statistics.

Table 3 - Tasks performed during internship.

Tasks	Nº of Tasks
Protocol Data	4
Protocol Data QC	5
Randomization Specification	4
Randomization Specification QC	3
Randomization List	4
Randomization List QC	3
Table of results	5
Table of results QC	5
Annotated CRF	1
Annotated CRF QC	1
SDTM Datasets	2
SDTM Datasets QC	1
ADaM Datasets	2
ADaM Datasets QC	2
TFLs	1
TFLs QC	1
SAP Review	3
Archive	3

6. Conclusion

In this report, the goals of providing an overview of the best practices and guidelines in the programming of TFLs for regulatory submission of Phase I clinical trials and exemplifying specific procedures of the creation of TFLs using the SAS programming language were achieved.

As a CDP, it is our job to create TFLs so that data can be presented in the CSR at the variable levels of detail, in the respective sections. It is also important to make that the TFLs follow a certain format in order for them to follow regulatory guidelines proposed by agencies such as FDA and EMA.

Regarding this internship as a CDP at BlueClinical, it helped me refresh some of the previous knowledge gained during the Master, as well as it made me acquire both Hard and Soft Skills that are needed in CDP. It also gave me an input on other Clinical Research procedures that are linked to other teams apart from CDP. At last, this internship not only gave me a new perspective on what it is like to work as a programmer in clinical research, but also gave me another perspective on the field of clinical research as a whole.

In conclusion, this curricular internship was a big asset for me to be able to integrate myself in the job market in the clinical research field.

7. Bibliography

1. BlueClinical, Ltd. About Us. Blue Clinical, Ltd. Accessed March 10, 2023. <https://www.blueclinical.com/about-us>.
2. BlueClinical, Ltd. Quality Manual V11.0. Published October 2022. Accessed March 10, 2023.
3. International Conference On Harmonisation Of Technical Requirements For Registration Of Pharmaceuticals For Human Use. E9 Statistical Principles for Clinical. FDA. Accessed March 15, 2023. <https://www.fda.gov/media/71336/download>. Published September 1998.
4. Clinical Data Interchange Standards Consortium. SDTM. CDISC. Accessed March 16, 2023. <https://www.cdisc.org/standards/foundational/sdtm>.
5. Clinical Data Interchange Standards Consortium. ADaM. CDISC. Accessed March 16, 2023. <https://www.cdisc.org/standards/foundational/adam>.

6. International Conference On Harmonisation Of Technical Requirements For Registration Of Pharmaceuticals For Human Use. ICH M2 EWG Electronic Common Technical Document Specification. ICH.org. Published July 16, 2008. Accessed March 18, 2023. https://admin.ich.org/sites/default/files/inline-files/eCTD_Specification_v3_2_2_0.pdf.
7. National Library of Medicine. Learn About Studies. ClinicalTrials.gov. Accessed April 20, 2023. <https://beta.clinicaltrials.gov/study-basics/learn-about-studies>.
8. U.S. Food and Drug Administration. What Are the Different Types of Clinical Research?. FDA. Published April 1, 2018. Accessed April 19, 2023. <https://www.fda.gov/patients/clinical-trials-what-patients-need-know/what-are-different-types-clinical-research>.
9. European Medicine Agency. Clinical Trials Regulation. EMA. Published February 23, 2023. Accessed April 21, 2023. <https://www.ema.europa.eu/en/human-regulatory/research-development/clinical-trials/clinical-trials-regulation>.
10. Assembleia da República. Lei n.º 21/2014. Diário da República Eletrónico. Published April 16, 2014. Accessed April 20, 2023. <https://dre.pt/dre/detalhe/lei/21-2014-25344024>.
11. Infarmed. Indicadores 2005 to 4Q2022. Infarmed. Published February 16, 2023. Accessed May 3, 2023. https://www.infarmed.pt/web/infarmed/servicos-online?p_p_id=101&p_p_lifecycle=0&p_p_state=maximized&p_p_mode=view&_101_struts_action=%2Fasset_publisher%2Fview_content&_101_returnToFullPageURL=%2Fweb%2Finfarmed%2Fservicos-online&_101_assetEntryId=8589127&_101_type=document&inheritRedirect=false&redirect=https%3A%2F%2Fwww.infarmed.pt%2Fweb%2Finfarmed%2Fservicos-online%3Fp_p_id%3D3%26p_p_lifecycle%3D0%26p_p_state%3Dmaximized%26p_p_mode%3Dview%26_3_redirect%3D%252Fweb%252Finfarmed%252Fservicos-online%26_3_paginationPhase%3Dtrue%26_3_keywords%3Dmedicamentos%252Bn%25EF%25BF%25BD%25EF%25BF%25BDo%252Bsujeitos%252Ba%252Breceita%26_3_advancedSearch%3Dfalse%26_3_groupId%3D15786%26_3_delta%3D30%26_3_assetTagNames%3Dindicadores%2Bde%2Batividade%26_3_resetCur%3Dfalse%26_3_andOperator%3Dtrue%26_3_struts_action%3D%252Fsearch%252Fsearch%26_3_reorderBy%3DorderByDate.
12. DataFlair. Advantages of SAS | Disadvantages of SAS programming. DataFlair. Published May 9, 2021. Accessed May 8, 2023. <https://data-flair.training/blogs/disadvantages-and-advantages-of-sas/>.

13. U.S. Food and Drug Administration. Study Data Technical Conformance Guide. FDA. Published March 2023. Accessed May 3, 2023. <https://www.fda.gov/media/153632/download>.
14. International Conference On Harmonisation Of Technical Requirements For Registration Of Pharmaceuticals For Human Use. ICH Official Website. ICH. Accessed May 10, 2023. <https://ich.org/page/mission>.
15. International Conference On Harmonisation Of Technical Requirements For Registration Of Pharmaceuticals For Human Use. Guideline for good clinical practice E6(R2). EMA. Published December 1, 2016. Accessed March 15, 2023. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-guideline-good-clinical-practice-e6r2-step-5_en.pdf.
16. U.S. Food and Drug Administration. What does FDA do?. FDA. Accessed March 15, 2023. <https://www.fda.gov/about-fda/fda-basics/what-does-fda-do>.
17. U.S. Food and Drug Administration. The Drug Development Process. FDA. Accessed March 15, 2023. <https://www.fda.gov/patients/drug-development-process>.
18. Code of Federal Regulations. Title 21 - Food and Drugs. ecf.gov. Published March 10, 2023. Accessed March 15, 2023. <https://www.ecfr.gov/current/title-21>.
19. Center for Drug Evaluation and Research. Computerized systems used in clinical investigational investigations. FDA. Published May 2007. Accessed March 15, 2023. <https://www.fda.gov/media/70970/download>
20. Center for Drug Evaluation and Research. Electronic source data in clinical investigations. FDA. Published September 2013. Accessed March 15, 2023. <https://www.fda.gov/media/85183/download>.
21. Center for Drug Evaluation and Research. Providing regulatory submissions in electronic format – standardized study data. FDA. Published June 2021. Accessed March 15, 2023. <https://www.fda.gov/media/82716/download>.
22. European Medicines Agency. Clinical trials in human medicines. EMA. Published March 9, 2023. Accessed March 15, 2023. <https://www.ema.europa.eu/en/human-regulatory/research-development/clinical-trials-human-medicines>.

23. European Medicines Agency. Clinical trials regulation. EMA. Published February 23, 2023. Accessed March 15, 2023. <https://www.ema.europa.eu/en/human-regulatory/research-development/clinical-trials/clinical-trials-regulation>.
24. European Medicines Agency. Covid-19 vaccines: Development, evaluation, approval and monitoring. EMA. Published March 3, 2023. Accessed March 15, 2023. <https://www.ema.europa.eu/en/human-regulatory/overview/public-health-threats/coronavirus-disease-covid-19/treatments-vaccines/vaccines-covid-19/covid-19-vaccines-development-evaluation-approval-monitoring>.
25. International Conference On Harmonisation Of Technical Requirements For Registration Of Pharmaceuticals For Human Use. ICH guideline M4 (R4) on common technical document (CTD) for the registration of pharmaceuticals for human use - organisation of CTD. ICH. Published March 19, 2021. Accessed May 3, 2023. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-guideline-m4-r4-common-technical-document-ctd-registration-pharmaceuticals-human-use_en.pdf.
26. European Medicines Agency. Good pharmacovigilance practices. EMA. Published February 10, 2023. Accessed March 15, 2023. <https://www.ema.europa.eu/en/human-regulatory/post-authorisation/pharmacovigilance/good-pharmacovigilance-practices>.
27. European Medicines Agency. Guideline on Data Monitoring Committee. EMA. Published July 27, 2005. Accessed March 20, 2023. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-data-monitoring-committees_en.pdf.
28. European Medicines Agency. Guideline on computerised systems and electronic data in clinical trials. EMA. Published March 9, 2023. Accessed March 15, 2023. https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/guideline-computerised-systems-electronic-data-clinical-trials_en.pdf.
29. Comissão de Ética para a Investigação Clínica. Missão e Visão. CEIC. Accessed March 19, 2023. <https://www.ceic.pt/missao>.
30. European Union. GDPR Archives. GDPR.eu. Published May 25, 2018. Accessed March 19, 2023. <https://gdpr.eu/tag/gdpr/>.

31. Assembleia da República. Lei n.º 58/2019, de 8 de agosto. Diário da República Eletrónico. Published August 8, 2019. Accessed March 19, 2023. <https://dre.pt/dre/detalhe/lei/58-2019-123815982>.
32. Infarmed. Apresentação. Infarmed. Accessed March 19, 2023. <https://www.infarmed.pt/web/infarmed/apresentacao>.
33. International Conference On Harmonisation Of Technical Requirements For Registration Of Pharmaceuticals For Human Use. E3 structure and content of clinical study reports. EMA. Published July 1996. Accessed May 11, 2023. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-3-structure-content-clinical-study-reports-step-5_en.pdf.
34. Clinical Data Interchange Standards Consortium. Standards. CDISC. Published 2023. Accessed April 12, 2023. <https://www.cdisc.org/standards>.
35. CDISC Submission Data Standards Team and CDISC SDTM Governance Committee. Study Data Tabulation Model v1.4. CDISC. Published November 26, 2013. Accessed April 12, 2023. <https://www.cdisc.org/standards/foundational/sdtm/sdtm-v1-7/html>.
36. CDISC Submission Data Standards Team. Study Data Tabulation Model Implementation Guide v3.2. CDISC. Published November 26, 2013. Accessed April 12, 2023. <https://www.cdisc.org/standards/foundational/sdtmig/sdtmig-v3-3/html>.
37. CDISC Analysis Data Model Team. Analysis Data Model (ADaM) Version 2.1. CDISC. Published November 29, 2021. Accessed April 13, 2023. <https://www.cdisc.org/standards/foundational/adam/adam-v2-1>.
38. CDISC Analysis Data Model Team. Analysis Data Model Implementation Guide v1.3. CDISC. Published November 29, 2021. Accessed April 13, 2023. <https://www.cdisc.org/standards/foundational/adam/adamig-v1-1>.
39. CDISC Analysis Data Model Team. ADaM structure for occurrence data (OCCDS) v1.0. CDISC. Published February 12, 2016. Accessed April 13, 2023. <https://www.cdisc.org/standards/foundational/adam/adam-structure-occurrence-data-occds-v1-0>.
40. Mahesh I. Visualizing Clinical Data with TLFS: Practical guide. ClinicalCrunch. Published April 19, 2023. Accessed May 11, 2023. <https://clinicalcrunch.com/tlfs-tables-listings-and-figures/>.

41. Chowdhury S. Generating tables, listings and figures without any programs. lexjansen. Published 2018. Accessed May 8, 2023. <https://www.lexjansen.com/phuse-us/2018/tt/TT09>.