

IDENTIFYING OUTLIER DATA POINTS IN NON-CLINICAL INVESTIGATIONAL NEW
DRUG SUBMISSIONS

By

Cassandra O'Malley, B.S.

A thesis submitted to the Graduate Committee of
Ramapo College of New Jersey in partial fulfillment
of the requirements for the degree of
Master of Science in Data Science
Spring, 2023

Committee Members:

Osei Tweneboah, Advisor

Scott Frees, Reader

Brenda Finney, Reader

COPYRIGHT

© Cassandra O'Malley

2023

Acknowledgements

This thesis has become formalized from initial thought to paper with the support of many individuals across a variety of academic, professional, and personal relationships, all of whom have encouraged me throughout this journey – thank you all.

I would like to thank following individuals for their specific roles in the submission of this study:

My advisor, Osei Tweneboah, for his guidance, reassurance, and commitment to success.

Reader Scott Frees, for his encouragement and counsel.

Reader Brenda Finney, for her wealth of knowledge in this field of study.

Michael Wasko, for his mentorship, support, introduction to the world of non-clinical/preclinical data, and motivation to complete the M.S. Data Science program.

Table of Contents

ACKNOWLEDGEMENTS	III
TABLE OF CONTENTS	IV
LIST OF TABLES	V
LIST OF FIGURES	VI
ABSTRACT	1
INTRODUCTION	2
BACKGROUND	5
METHODOLOGY	9
OVERVIEW OF SEND VARIABLES	10
TX – TRIAL SETS.....	11
DM – DEMOGRAPHICS	12
BW – BODY WEIGHTS.....	13
<i>Change from Baseline Flag</i>	14
<i>Changes from Control Group (Percent)</i>	15
LB – LABORATORY	15
<i>Changes from Baseline Flag</i>	16
<i>Changes from Control Group (Percent)</i>	17
<i>Changes from Group (Standard Deviation)</i>	17
OM – ORGAN MEASUREMENTS.....	18
<i>Changes from Group (Standard Deviation)</i>	18
MA – MACROSCOPIC FINDINGS.....	19
<i>Unremarkable vs. Remarkable Findings</i>	20
MI – MICROSCOPIC FINDINGS	20
<i>Unremarkable vs. Remarkable Findings</i>	21
TEST DATA.....	21
<i>DART Fit-For-Use Pilot</i>	21
<i>TEST1 Dataset</i>	22
ANALYSIS AND DISCUSSION	25
CONCLUSIONS	29
REFERENCES	32
APPENDIX A – DEFINITIONS	35

List of Tables

Table 4.1: Example of TX domain with information regarding the presence of the control type	11
Table 4.2: Example of DM domain with information to the sex of the subject and the set	12
Table 4.3: Example of BW domain for two subjects from start of the study to their day of removal from the study	14
Table 4.4: Example of LB domain for three test taken on two different days.....	16
Table 4.5: Example of OM domain for two animals with different specimens.....	18
Table 4.6: Example of MA domain with a remarkable finding for one specimen	19
Table 4.7: Example of MI domain with positive findings and unremarkable specimens.....	20
Table 5.1: Bodyweight findings for three subjects in the BW domain with methods listed	26
Table 5.2: Single subject findings for the BW and OM domain	26

List of Figures

Figure 3.1: Drug development process from research initiation to final approval (Hughes et al., 2011)	6
Figure 4.1: Body weight means of TEST1 study by a group with 90-110% of control group weight highlighted	23
Figure 4.2: Selected organ weight means by group in the study TEST1	24
Figure 5.1: Remarkable and unremarkable findings in a group of animals; specimens were only noted individually if there were findings	27

Abstract

The Food and Drug Administration (FDA) uses a format known as SEND (Standard for Exchange of Nonclinical Data) to evaluate non-clinical (animal) studies for investigational new drug applications. Investigative drug sponsors currently use information from historical and control data to determine if drugs cause toxicity.

The goal of this study is to identify outlying data points that may indicate an investigational new drug could be toxic. Examples include a negative body weight gain over time, enlarged organ weights, or laboratory test abnormalities, especially in relation to a control group within the same study. Flagged records can be analyzed by a veterinarian or pathologist for potential signs of toxicity without looking at each individual data point.

Common domains within the non-clinical pharmaceutical studies were evaluated using changes from baseline measurements, changes from the control group, a percent change from the previous measurement with reference to the ethical guidelines, values outside of the mean \pm two standard deviations, and a measure of abnormal findings to unremarkable findings in pathology. A program was designed to analyze five of these domains and return a collection of possible outlying data for simpler and faster than individual data point analysis by a study monitor, performing the analysis in a fraction of the time. The resulting file is more easily read by someone unfamiliar with the SEND format.

With this program, analyzing a study for possible toxic effects during the study can save time, effort, and even animal lives by identifying the signs of toxicity early. Sponsors or CROs can determine if the product is safe enough to proceed with testing or should be stopped in the interest of safety and additional research.

Introduction

The Food and Drug Administration (FDA) is responsible for regulating items such as food, drugs, medical devices, vaccines, and biologics within the United States, protecting the public and ensuring the nation's safety. The FDA's Center for Drug Evaluation and Research (CDER) monitors the safety and effectiveness of proposed and approved drugs. Within CDER, certain information submission practices are standardized regarding guidance, compliance, and regulation. To submit new drugs for approval to the FDA, protocols, and studies are submitted to the FDA for approval and interim monitoring.

Two of the most common applications to the FDA are Investigational New Drug applications (IND) and New Drug Applications (NDA) (FDA, 2014); the difference between an IND and an NDA is that the IND must be approved to begin testing past the most basic safety assessments (non-clinical studies), versus an NDA is seeking approval at the end of clinical (human) testing. In non-clinical studies submitted for IND approval, the data discovered in the research phases must be submitted to the FDA in a specific format – this is known as the Standard for Exchange of Nonclinical Data (SEND). The study performed for a submission has endpoints that can be quantified and qualified in specific measures, which can be expressed in this SEND dataset. The FDA, in turn, uses this SEND dataset to determine if the IND application is safe for humans by reviewing the investigational drug at several dose levels, using various methods, and in different species for adverse reactions.

Each of these studies must be monitored by several personnel to ensure the validity of the research and its' results, compliance with regulatory authorities, and maintaining the humane endpoints as set forth by the governing bodies. With many studies in progress across a test

facility, a single veterinarian, study director, or study monitor, could be overwhelmed, especially with large studies – carcinogenicity studies, for example, can have more than 500 subjects in a study and last for over two years. By identifying signs of toxicity early, an investigator can be alerted to potential problems early and remove a subject from treatment or administer care before there is undue suffering.

The purpose of the study presented in this thesis is to produce a deliverable that can interpret the data in the SEND format, and flag outlying data points compared to the rest of the performed study. A study monitor person can evaluate if the outlying data is a sign of toxicity by identifying an individual with a much larger or smaller value than the control group or the rest of the group to which the subject is assigned. For numerical findings, this will be done by statistical analysis, split out in the study according to sex, dose levels, and different study days to allow for a growth rate. For textual findings, this will be done by a number of remarkable findings compared to the unremarkable or “normal” findings; some remarkable findings are common per species, while others are a sign of toxicity.

The information identified is to be used as a potential warning sign for the study monitoring personnel and is not to be used for diagnosis. A study staff member should understand that this tool is utilized for an extremely quick overview of quantitative anomalies instead of manually reviewing each data point.

A background into the SEND dataset and preclinical/non-clinical drug trials are included in this document. Humane endpoints given by the regulatory agencies are outlined and serve as the basis of analysis criteria. Specific mentions are made towards the individual pieces of data that

can be analyzed through this program, as well as the ideas towards development to include more domains and standards as future work.

In the background of this document, the history of the drug development process for approval by the FDA is described, starting in the early 1900s with the need for studies to prevent poisoning. Research on drug development takes upwards of 10-15 years and billions of dollars to successfully make it to market, whereas the same amount of effort and funding is put into drugs that are eventually failed along the same pathway. Efforts made in this study will attempt to reduce the resources put into studies that may or may not ultimately be approved.

The SEND data that is analyzed in this study is described in the methodology section of this document. Specific portions of the SEND requirement were utilized where the control group or baseline flag could be used as a reference point. The data used to test the program was included, both publicly available and created for the purpose of this study.

Background

The control of food and medicine safety dates back to the first food law in 1202 by King John of England, which defined the standards of bread, prohibiting additive ingredients (FDA, 2023). The first biologic law in the United States was passed in 1813 regarding the smallpox vaccine, and the federal control of food and drugs began in 1848. In 1906, a similar law to the 1202 bread standard was passed prohibiting contaminated food, drink, and drugs.

In 1937, the S.E. Massengill Co. developed an oral drug named “Elixir Sulfanilamide” (Ballentine, 1981). The drug sulfanilamide was used in tablet and powder form for streptococcal infections (Hajar, 2011), and the S.E. Massengill Co. found that the compound would dissolve in the substance diethylene glycol, to create an oral dosage form. As the tablet and powder forms were previously known to treat infections successfully, the newly discovered oral form was quickly compounded with a few other substances to improve flavor and appearance of the product and shipped across the United States. Within approximately one month of the first shipment, at least one hundred patients died from what we now know to be a poisonous substance – diethylene glycol is chemically similar to antifreeze (ethylene glycol), which is toxic to mammals; the chemical in even small amounts can cause kidney failure, leading to permanent injury, or death. The 1937 poisonings and deaths spurred action by the Food and Drug Administration, enacting the 1938 Federal Food, Drug, and Cosmetic Act. The act mandated that all new drugs be proven safe and effective, which began the widespread testing of new substances on animals.

In 1962, as a response to the thalidomide birth defect crisis (Vargesson, 2015), the FDA introduced the “Kefauver-Harris Drug Amendments”, a set of specifications to amend the

previous act and establish the rules of investigation of new drugs (Meadows, 2006). Amid the AIDS epidemic, the first formal establishment of an IND and clinical trials began with the “Treatment IND” in 1987 (White Junod, 2008).

Further development of these regulations resulted in the development of the “Investigational New Drug” application (IND) in 1992, the CDER guidance of 1999, and the Study Data Tabulation Model (SDTM) for clinical studies (Wood, 2011). Through the early 2000s, a small sub-team (industry and FDA) worked to develop the Standard for the Exchange of Nonclinical Data (SEND) for nonclinical data. The purpose of the SEND model is to structure, organize, and format the nonclinical data for submission and review in a singular format (Charles River). Reviewing paper applications takes an immense amount of time; by developing a standard for submission electronically, statistical analysis is driven down to minutes, and initial review of data can take a few hours, as opposed to days or weeks, of reviewing individual pieces of paper. Creating a standard for exchanging nonclinical information has many benefits for both the sponsor organization in time and effort saved for paper submissions, and for the FDA in reducing the time to review.

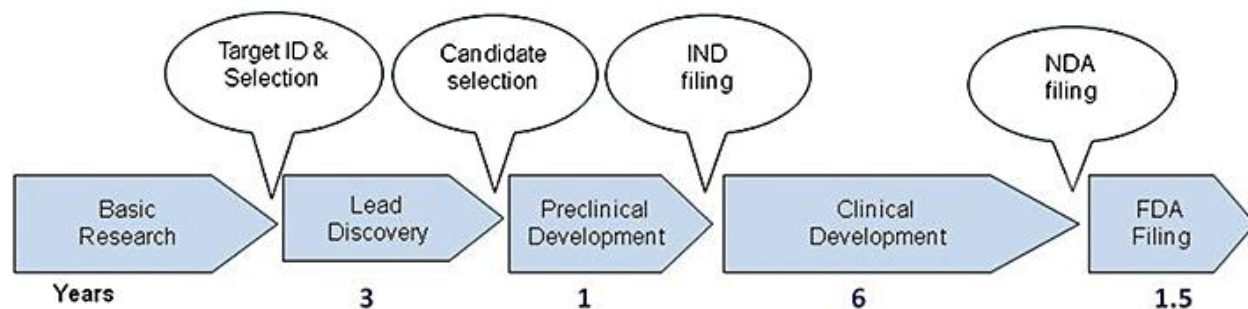


Figure 3.1: Drug development process from research initiation to final approval (Hughes et al., 2011)

Figure 3.1 displays an overview of the drug development process with an approximate timeline of 11.5 years of testing. This excludes the initial basic research timeline which is spent typically in exploration and development and can last for many years. The current drug development process takes approximately 10 to 15 years and costs several millions of dollars. One review of submissions estimates that the actual cost of a drug from initiation to the market is over \$1 billion (Van Norman, 2016). Only a fraction of drugs that are submitted to the FDA moves onto each step, with an estimate that only 10% of applications make it through each phase, or a 90% failure rate at each phase from preclinical to phase III trials (Sun, 2022). This does not account for the test articles that are failed out of the process in the basic research step, which could be thousands before a successful candidate is found to begin the discovery process.

With a small amount of drugs succeeding through each phase, an organization would want to minimize their investment into the drugs that will not be approved at the next stage. By failing early, a company can minimize the effects of failure of a test article by reducing the time, effort, and cost associated with trying to get a drug to market. By efficiently analyzing the effects of a test article in the preclinical phase before submission, a team might be able to get ahead on the outcome of the FDA's review – if a test article is seen in preclinical tests as unacceptably toxic, the sponsor of the drug may choose to end the research on the product earlier than expected, saving not only a significant amount of time and money but also the lives of potentially hundreds of animals per test.

The good laboratory practice (GLP) guidelines around preclinical studies and their related legislations dictate humane endpoints for animal studies, indicating distress, pain, or suffering. Typically, this indicates a level of toxicity in the treatment. By waiting for an observation of this clear distress, the animal may already be suffering; to prevent this suffering,

statistical analyses can be utilized throughout the study to find outlying data points that may indicate an animal is falling outside of a range that is considered normal, or if an entire group within the study is showing signs of being overdosed. With this deliverable, I aim to prove that an automated statistical analysis program on already collected data can identify signs of humane endpoints early and prevent undue stress and suffering.

Methodology

In an effort to visualize the outlying data, it must be understood how the data can be presented and processed, as well as what the data represents. The term “dataset” will represent the data from a singular preclinical study from study initiation until final signature, as organized into SEND 3.1 standards, which is the current standard accepted by the FDA (Center for Drug Evaluation and Research, 2017). The SEND Implementation Guide (SENDIG) defines the types of data that can be represented in a singular “domain”, a file representing a subset of data within the data. A two-letter abbreviation denotes each domain, i.e., “BW” stands for “Body Weights”, and “EX” for “Exposure”. Domains are categorized into subsections: special purpose, interventions, events, findings, trial design, and relationship domains. Additionally, each of the domains also has a reference to what kind of data the domain holds, i.e., whether the data is recorded in-life or is post-mortem. The analysis will focus on in-life and post-mortem domains, all categorized under “Findings”. Metadata will need to be retrieved from the trial design information, but the trial design information does not contain any information for data point analysis.

The standard format to submit SEND datasets to the FDA is in SAS transport files, with the file extension of “.xpt”. The format was designed to hold data for short-term transfer within statistical software systems (SAS Transport File Format (XPT) Family); a freely available software to visualize the software is available through the SAS Institute. As datasets are transported within the .xpt format for formal approval, a program to visualize the information from the largest amount of data sources would involve reading the .xpt format. All laboratory information management systems (LIMS) record, store, and export data in different formats,

therefore the .xpt extension harmonizes data from a large majority of the different types of recording systems. As paper-recorded items also have to be transposed into the .xpt format for submission, it is the goal that the tool developed can read 100% of the datasets that will eventually be submitted under the categories that require SEND, with the caveat that if a dataset that does not have findings domains, there are no data points to analyze.

For the amounts of data that could be processed, Python was chosen to load most of the information. The Python program, herein referred to as the “analysis program”, “program”, and “tool”, was designed to look for .xpt formats within a specified pathway with a folder at the same location as the program. The pandas library was included to read in the .xpt format, and will further run analysis techniques. An additional, optional step of translating .xls/.xlsx documents to .xpt format was created in R so that all datasets can be processed in the data frames created from .xpt format.

The goal at the end of the study is to produce a deliverable that can clearly identify which records are out of the acceptable range according to other records within the study. Each record in the deliverable document should include the subject number, test, study day (if applicable), specimen (if applicable), and, most importantly, how the outlying status was assigned to the record through statistical analysis.

Overview of SEND variables

The SEND requirement has a few terms that will be used repeatedly, across each dataset. The “STUDYID” parameter represents a unique study identifier that is going to be used to represent the study in a submission of multiple studies with the same test article identifier. The “DOMAIN” parameter is a two-letter abbreviation of the domain. The abbreviation “—SEQ” is a

sequence number that is unique to the domain or the subject, depending on the domain it is used in. The variable “USUBJID” stands for “Unique Subject Identifier” – this is a combination of the STUDYID, hyphenated with the subject identifier (number) that is usually presented in the study report. This creates a unique identifier to refer to the subject across not only the study, but across the entire submission of the test article.

TX – Trial Sets

One of the concepts of testing an investigative treatment is the presence of a “control group”, a group that does not receive the investigative treatment and receives either no treatment, or a controlled treatment such as a vehicle. One example of a treated versus control group would be a subject receiving an injection of a test article, but the control group subjects receive an injection of phosphate buffered saline, which is commonly a dilutant and vehicle to transport test articles and is isotonic to the body.

STUDYID	DOMAIN	SETCD	SET	TXSEQ	TXPARMCD	TXPARAM	TXVAL
TEST1	TX	1	0 mg/kg TRT	1	ARMCD	Arm Code	1
TEST1	TX	1	0 mg/kg TRT	2	SPGRPCD	Sponsor-Defined Group Code	1
TEST1	TX	1	0 mg/kg TRT	3	GRPLBL	Group Label	0 mg/kg
TEST1	TX	1	0 mg/kg TRT	4	TRTDOS	Dose Level	0
TEST1	TX	1	0 mg/kg TRT	5	TRTDOSU	Dose Units	mg/kg
TEST1	TX	1	0 mg/kg TRT	6	TCNTRL	Control Type	Vehicle Control
TEST1	TX	1	0 mg/kg TRT	7	PLANMSUB	Planned Number of Male Subjects	3
TEST1	TX	1	0 mg/kg TRT	8	PLANFSUB	Planned Number of Female Subjects	3
TEST1	TX	1	0 mg/kg TRT	9	SPLANSUB	Planned Number of Subjects	6
TEST1	TX	2	10 mg/kg TRT	10	ARMCD	Arm Code	2
TEST1	TX	2	10 mg/kg TRT	11	SPGRPCD	Sponsor-Defined Group Code	2
TEST1	TX	2	10 mg/kg TRT	12	GRPLBL	Group Label	10 mg/kg
TEST1	TX	2	10 mg/kg TRT	13	TRTDOS	Dose Level	10
TEST1	TX	2	10 mg/kg TRT	14	TRTDOSU	Dose Units	mg/kg
TEST1	TX	2	10 mg/kg TRT	15	PLANMSUB	Planned Number of Male Subjects	3
TEST1	TX	2	10 mg/kg TRT	16	PLANFSUB	Planned Number of Female Subjects	3
TEST1	TX	2	10 mg/kg TRT	17	SPLANSUB	Planned Number of Subjects	6

Table 4.1: Example of TX domain with information regarding the presence of the control type

Table 4.1 displays an example of the TX domain, which represents different characteristics of a group of treated animals. The columns “SETCD” and “SET” represent the different codes/names for the group, and the associated variables of the group are on the right,

represented with “TXPARAMCD” as the parameter code, “TXPARAM” as the long name for the parameter, and the “TXVAL” which is the value for that group’s parameter. As an example, group 1 has a total of 6 animals in the group, 3 males and 3 females, represented by lines 7-9 (TXSEQ 7-9).

The TX domain contains information on if each group tested belongs to a control group, which can help to indicate if the effect of the treatment is known or can be deduced from the lack of treatment. If the group contains the parameter “Control Type”, the group number will be recorded to be used as a control group in analysis, a form of baseline to compare results against.

DM – Demographics

The purpose of the demographics domain is to define the variable information about each subject in the study. This will contain not only each subject’s identifier but age, sex, group assignment, and start and end dates of treatment.

STUDYID	DOMAIN	USUBJID	SUBJID	RFSTDTDC	RFENDTDC	RFXSTDTDC	RFXENDTDC	AGE	AGEU	SEX	SPECIES	STRAIN	ARMCD	ARM	SETCD
TEST1	DM	TEST1-1001	1001	2023-01-01	2023-01-29	2023-01-01	2023-01-28	14	MONTHS	M	DOG	BEAGLE		1 0 mg/kg TRT	1
TEST1	DM	TEST1-1002	1002	2023-01-01	2023-01-29	2023-01-01	2023-01-28	16	MONTHS	M	DOG	BEAGLE		1 0 mg/kg TRT	1
TEST1	DM	TEST1-1003	1003	2023-01-01	2023-01-29	2023-01-01	2023-01-28	15	MONTHS	M	DOG	BEAGLE		1 0 mg/kg TRT	1
TEST1	DM	TEST1-1004	1004	2023-01-01	2023-02-12	2023-01-01	2023-01-28	14	MONTHS	M	DOG	BEAGLE	1R	0 mg/kg TRT, Recovery	1R
TEST1	DM	TEST1-1005	1005	2023-01-01	2023-02-12	2023-01-01	2023-01-28	15	MONTHS	M	DOG	BEAGLE	1R	0 mg/kg TRT, Recovery	1R
TEST1	DM	TEST1-1501	1501	2023-01-02	2023-01-30	2023-01-02	2023-01-29	14	MONTHS	F	DOG	BEAGLE		1 0 mg/kg TRT	1
TEST1	DM	TEST1-1502	1502	2023-01-02	2023-01-30	2023-01-02	2023-01-29	16	MONTHS	F	DOG	BEAGLE		1 0 mg/kg TRT	1
TEST1	DM	TEST1-1503	1503	2023-01-02	2023-01-30	2023-01-02	2023-01-29	12	MONTHS	F	DOG	BEAGLE		1 0 mg/kg TRT	1
TEST1	DM	TEST1-1504	1504	2023-01-02	2023-02-13	2023-01-02	2023-01-29	13	MONTHS	F	DOG	BEAGLE	1R	0 mg/kg TRT, Recovery	1R
TEST1	DM	TEST1-1505	1505	2023-01-02	2023-02-13	2023-01-02	2023-01-29	15	MONTHS	F	DOG	BEAGLE	1R	0 mg/kg TRT, Recovery	1R
TEST1	DM	TEST1-2001	2001	2023-01-01	2023-01-29	2023-01-01	2023-01-28	12	MONTHS	M	DOG	BEAGLE		2 10 mg/kg TRT	2
TEST1	DM	TEST1-2002	2002	2023-01-01	2023-01-29	2023-01-01	2023-01-28	14	MONTHS	M	DOG	BEAGLE		2 10 mg/kg TRT	2
TEST1	DM	TEST1-2003	2003	2023-01-01	2023-01-29	2023-01-01	2023-01-28	16	MONTHS	M	DOG	BEAGLE		2 10 mg/kg TRT	2
TEST1	DM	TEST1-2501	2501	2023-01-02	2023-01-30	2023-01-02	2023-01-29	13	MONTHS	F	DOG	BEAGLE		2 10 mg/kg TRT	2
TEST1	DM	TEST1-2502	2502	2023-01-02	2023-01-30	2023-01-02	2023-01-29	12	MONTHS	F	DOG	BEAGLE		2 10 mg/kg TRT	2
TEST1	DM	TEST1-2503	2503	2023-01-02	2023-01-30	2023-01-02	2023-01-29	12	MONTHS	F	DOG	BEAGLE		2 10 mg/kg TRT	2

Table 4.2: Example of DM domain with information to the sex of the subject and the set

In Table 4.2, the DM domain details information such as the start and end date of the treatment (RFXSTDTDC and RFXENDTDC), and the subject’s start and end date on the study (RFSTDTDC and RFENDTDC). The subject’s each have an associated age represented by AGE and AGEU (unit), and have their sex recorded as male or female represented as “M” or “F” in the column SEX. The species and strain of each animal are recorded, which will typically be the

same value for the entire study; it is extremely rare to see different strains of animals within the same study, and practically impossible to use different species of animals in the same study. The last 3 columns, ARMCD, ARM, and SETCD represent the different paths of treatment each subject takes. It can be noted that subjects with an “R” in ARMCD or “, Recovery” in ARM are part of a recovery group; these subjects receive a two-week non-treatment period after the four-week treatment period to monitor recovery from the side effects of the test article.

The demographics domains is used in this program to add metadata to each findings domain for use in splitting results by sex and group number. Biologically, different sexes should not be compared. Additionally, the group information that is found in DM and TX can be added to the other domains to look at how different dose levels may impact changes on individual and group statistics.

BW – Body Weights

The BW domain includes every body weight recorded for each animal; there is typically a pre-study weight taken before dosing, body weights recorded at set intervals (i.e., one time per week), and terminal body weights recorded at the time of sacrifice during the study. Each record contains information as to the subject identifier, the weight, and the day of recording; study day 1 corresponds to the first day of dosing for that subject on the study. As this is the most basic findings domain, having only two types of tests (body weight, terminal body weight), and all records have a numeric finding, analysis was developed first on the body weight domain.

STUDYID	DOMAIN	USUBJID	BWSEQ	BWTESTCD	BWTEST	BWORRES	BWORRESU	BWSTRESC	BWSTRESN	BWSTRESU	BWBLFL	BWFAST	BWDTC	BWDY	BWNOMDY	BWNOMLBL
TEST1	BW	TEST1-1001	1	BW	Body Weight	11.37		11.37	11.37	kg			2022-12-24	-8		-8 Day -8
TEST1	BW	TEST1-2001	11	BW	Body Weight	11.64		11.64	11.64	kg			2022-12-24	-8		-8 Day -8
TEST1	BW	TEST1-1001	33	BW	Body Weight	10.34		10.34	10.34	kg	Y		2022-12-31	-1		-1 Day -1
TEST1	BW	TEST1-2001	43	BW	Body Weight	10.91		10.91	10.91	kg	Y		2022-12-31	-1		-1 Day -1
TEST1	BW	TEST1-1001	65	BW	Body Weight	9.67		9.67	9.67	kg			2023-01-01	1		1 Day 1
TEST1	BW	TEST1-2001	75	BW	Body Weight	10.32		10.32	10.32	kg			2023-01-01	1		1 Day 1
TEST1	BW	TEST1-1001	97	BW	Body Weight	9.19		9.19	9.19	kg			2023-01-08	8		8 Day 8
TEST1	BW	TEST1-2001	107	BW	Body Weight	11.24		11.24	11.24	kg			2023-01-08	8		8 Day 8
TEST1	BW	TEST1-1001	129	BW	Body Weight	9.65		9.65	9.65	kg			2023-01-15	15		15 Day 15
TEST1	BW	TEST1-2001	139	BW	Body Weight	11.54		11.54	11.54	kg			2023-01-15	15		15 Day 15
TEST1	BW	TEST1-1001	161	BW	Body Weight	9.15		9.15	9.15	kg			2023-01-22	22		22 Day 22
TEST1	BW	TEST1-2001	171	BW	Body Weight	12.35		12.35	12.35	kg			2023-01-22	22		22 Day 22
TEST1	BW	TEST1-1001	193	TERMBW	Terminal Body Weight	9.78		9.78	9.78	kg		Y	2023-01-29	29		29 Day 29
TEST1	BW	TEST1-2001	203	TERMBW	Terminal Body Weight	13.06		13.06	13.06	kg		Y	2023-01-29	29		29 Day 29

Table 4.3: Example of BW domain for two subjects from start of the study to their day of removal from the study

Table 4.3 gives an example of the body weight domain for two subjects, recording the body weights from the start of the study on day -8, through their scheduled removal from the study on day 29. The body weight fluctuates at each recording, with first a loss of body weight, but then the second subject, TEST1-2001, begins to gain body weight. On day -1 of the study, each animal has the column “BWBLFL” populated which means that their body weights on that day are to be considered the baseline weight for that animal; as this species is a dog, which is fully grown at the start of the study, rapid body weight loss or gain is not to be expected, and any sudden change can indicate a problem with the test article in the species/specific strain.

Change from Baseline Flag

In the BW domain, as well as a few other findings domains, there exists a parameter abbreviated as “--BLFL”, in which the first two letters designate the domain abbreviation, and “BLFL” stands for baseline flag. If the variable is populated as “Y”, the observation is considered a baseline observation for the subject.

In fully grown animals, especially adult primates, it is expected that body weights are stable, and there is little change that would be due to only environmental factors. A large change from the last body weight taken before dosing may indicate that the treatment could be impacting the body weight, or more concerning, that a loss of body weight is a sign of toxicity. By

identifying a large weight loss from the animal's weight at the start of the study, it is a sign that the study monitoring personnel should look for other signs that may be present as signs of toxicity. This particular assessment will point towards individual animals, but patterns of sustained weight loss across all animals may indicate the failure of a test article for toxicity in the test species.

Changes from Control Group (Percent)

By combining the body weight domain with the demographics information, and cross referencing the TX domain, the BW domain can be sub sectioned to each group. The control group, if identified in the TX analysis, can serve as the baseline observation in this analysis. The observations are split by sex and analyzed compared to the control group. The control group mean (per sex, per day) is calculated, and the animals of the same sex on the same day are compared to the control group's mean. If an animal has a body weight over a set difference by percent, the specific record is identified. The OECD recommends that body weights be identified for veterinary care, up to and including humane sacrifice, if the body weight is more than a 20% loss compared to the control group. The deliverable begins to identify body weights at a 10% loss so that early veterinary care may be actioned, if necessary, at that point in time.

LB – Laboratory

The laboratory domain contains information regarding laboratory tests, which can include but are not limited to hematology, coagulation, and clinical chemistry tests, typically performed on blood samples taken throughout the study. Examples of this include white blood cell count, and individual differentials, including both pure counts and ratios against the total amount of

white blood cells. Each record contains information about the subject, the test name, results and associated units, and the day of recording.

STUDYID	DOMAIN	USUBJID	LBSEQ	LBTESTCD	LBTEST	LBCAT	LBORRES	LBORRESU	LBSTRESC	LBSTRESN	LBSTRESU	LBSPEC	LBLOC	LBDMTC	LBDMY	LBDMOMDY
Study123	LB	Study123-19-0001	1	PT	Prothrombin Time	COAGULATION	7.6 sec		7.6	7.6 sec		PLASMA	Ear Artery	2019-02-01T07:31:40	19	19
Study123	LB	Study123-19-0001	2	FIBRINO	Fibrinogen	COAGULATION	261 mg/dL		261	261 mg/dL		PLASMA	Ear Artery	2019-02-01T07:31:40	19	19
Study123	LB	Study123-19-0001	3	APTT	Activated Partial Thromboplastin Time	COAGULATION	14.3 sec		14.3	14.3 sec		PLASMA	Ear Artery	2019-01-18T07:38:37	5	5
Study123	LB	Study123-19-0001	4	FIBRINO	Fibrinogen	COAGULATION	310 mg/dL		310	310 mg/dL		PLASMA	Ear Artery	2019-01-18T07:38:37	5	5
Study123	LB	Study123-19-0001	5	PT	Prothrombin Time	COAGULATION	7.4 sec		7.4	7.4 sec		PLASMA	Ear Artery	2019-01-18T07:38:37	5	5
Study123	LB	Study123-19-0001	6	APTT	Activated Partial Thromboplastin Time	COAGULATION	23.9 sec		23.9	23.9 sec		PLASMA	Ear Artery	2019-02-01T07:31:40	19	19

Table 4.4: Example of LB domain for three test taken on two different days

Table 4.4 displays information for subject number Study123-19-0001 on two different days – this table shows the tests Prothrombin Time, Fibrinogen, and Activated Partial Thromboplastin Time, all coagulation tests, on days 5 and 19. The tests were evaluated from the plasma collected from the ear artery of the subject. From day 5 to day 19, there was an increase in prothrombin time, a decrease in fibrinogen, and an increase in activated partial thromboplastin time. Across multiple recordings, the differences can be evaluated if they are within a normal limit, or if the changes indicate a sign of toxicity.

Changes from Baseline Flag

Though not as common as the BW domain, the LB domain contains the baseline flag parameter, noted as “LBBLFL”. If a record has a “Y” in LBBLFL, the data is to be used as a baseline for that animal. Typically, this is recorded in the last observation before dosing, either on day -1 or day 1. Baseline flags are usually only found in larger animals, such as primates, canines, and swine animals, as the blood collection processes from smaller animals are done at euthanasia due to the large amount of blood required compared to what the body can hold.

When the baseline flag is populated in the LB domain for a dataset, the DM domain is appended onto the LB domain for sub sectioning. The LB domain can be split into the groups differentiated by sex, group number, and is split by each test. A list is created to iterate through each test and compare each animal’s baseline flagged test against the further tests. For example, if male subject 1001 has a baseline flag on day -1 for the test Bilirubin, subject 1001’s other

Bilirubin tests on days 8, 22, and 43 are examined for a difference percentage. The record is flagged for inclusion into the outlier report if the difference is more than a specified percentage.

If there are significant differences across specific tests, i.e., all tests related to liver function, this may indicate that the test article impacts the liver function in the animal species it is being tested on. If there are increases or drops of more than an indicated range of acceptance, such as twenty or twenty-five percent, this could indicate a sign of toxicity. Each of these indicated changes would have to be examined by a scientist monitoring the study to evaluate the impact the test article may be having, as an outlying value may be due to other factors in the study that the analysis tool cannot understand, such as a different feed, a change in dosing, or a result of the environmental factors surrounding the study.

Changes from Control Group (Percent)

When combined with the DM domain, the LB domain can be iterated through by group. When a control group is indicated in the TX domain, the control group's code(s) is noted, and passed through into a function with the combined LB/DM domain. With the control group identified, the program separates the LB domain into different sexes, and then into the different control groups so that different sexes are not compared.

Each test is evaluated separately, as well as being assessed on separate days. The test article may impact the overall treatment, and that would be seen by the identification of many tests for different animals on the same days of the study. Each study monitor or study design would have its own cutoff, i.e., a change in 50% of the group.

Changes from Group (Standard Deviation)

Similar to the change in the control group using the percent difference, every animal can be compared to the rest of their group by the standard deviation. The LB domain is split amongst

the sexes and groups, and then every test and animal is evaluated. The mean and standard deviation are calculated, and the program identifies a value that is ± 2 standard deviations from the mean. If a record is flagged multiple times by these evaluations, it is a more likely indicator that the animal’s value is outlying and may be a sign of toxicity or other change in environment.

OM – Organ Measurements

The organ measurements (OM) domain represents information collected at necropsy, which includes the subject identifier, the weight (or other measurement parameter), result and associated unit, specimen, and the date taken. Most commonly, subjects that receive necropsy and have organ measurements taken, there will be a test of “WEIGHT”, and the weight of each specified organ is taken. Examples may include heart, lungs, and brain weights.

STUDYID	DOMAIN	USUBJID	OMSEQ	OMTESTCD	OMTEST	OMORRES	OMORRESU	OMSTRESC	OMSTRESN	OMSTRESU	OMSTAT	OMREASND	OMSPEC	OMDTC	OMDY	OMNOMDY	OMNOMLBL
TEST1	OM	TEST1-4001	23	WEIGHT	Weight	1.450	g	1.450	1.450	g			GLAND, ADR	2023-01-29	29	29	Day 29
TEST1	OM	TEST1-4002	24	WEIGHT	Weight						NOT DONE	Organ not weighed	GLAND, ADR	2023-01-29	29	29	Day 29
TEST1	OM	TEST1-4001	55	WEIGHT	Weight	33.45	g	33.45	33.45	g			BRAIN	2023-01-29	29	29	Day 29
TEST1	OM	TEST1-4002	56	WEIGHT	Weight	51.23	g	51.23	51.23	g			BRAIN	2023-01-29	29	29	Day 29
TEST1	OM	TEST1-4001	87	WEIGHT	Weight	80.29	g	80.29	80.29	g			HEART	2023-01-29	29	29	Day 29
TEST1	OM	TEST1-4002	88	WEIGHT	Weight	126.4	g	126.4	126.4	g			HEART	2023-01-29	29	29	Day 29
TEST1	OM	TEST1-4001	119	WEIGHT	Weight	20.07	g	20.07	20.07	g			KIDNEY	2023-01-29	29	29	Day 29
TEST1	OM	TEST1-4002	120	WEIGHT	Weight	48.95	g	48.95	48.95	g			KIDNEY	2023-01-29	29	29	Day 29
TEST1	OM	TEST1-4001	151	WEIGHT	Weight	390.3	g	390.3	390.3	g			LIVER	2023-01-29	29	29	Day 29
TEST1	OM	TEST1-4002	152	WEIGHT	Weight	375.7	g	375.7	375.7	g			LIVER	2023-01-29	29	29	Day 29
TEST1	OM	TEST1-4001	183	WEIGHT	Weight	92.55	g	92.55	92.55	g			LUNG	2023-01-29	29	29	Day 29
TEST1	OM	TEST1-4002	184	WEIGHT	Weight	54.64	g	54.64	54.64	g			LUNG	2023-01-29	29	29	Day 29
TEST1	OM	TEST1-4001	215	WEIGHT	Weight	34.57	g	34.57	34.57	g			PANCREAS	2023-01-29	29	29	Day 29
TEST1	OM	TEST1-4002	216	WEIGHT	Weight	40.98	g	40.98	40.98	g			PANCREAS	2023-01-29	29	29	Day 29
TEST1	OM	TEST1-4001	247	WEIGHT	Weight	55.87	g	55.87	55.87	g			SPLEEN	2023-01-29	29	29	Day 29
TEST1	OM	TEST1-4002	248	WEIGHT	Weight	41.66	g	41.66	41.66	g			SPLEEN	2023-01-29	29	29	Day 29

Table 4.5: Example of OM domain for two animals with different specimens

Table 4.5 displays two animals and their organ weights taken at necropsy. Subject TEST1-4002 has a “NOT DONE” records for the adrenal gland (OMSEQ 24) and will not be included in any analysis. The weights vary widely between the two subjects, which could be considered normal depending on their body weight, age, and sex; however, enlarged organ weights may indicate a sign of toxicity.

Changes from Group (Standard Deviation)

As organ weights vary by species, sex, and the age at the time of sacrifice, the mean and standard deviation are the most applicable method to finding the outlying organ weights. Sex and

specimens are split to run as a looping analysis, and each group is evaluated against other members of the group for each run. The mean and standard deviation are calculated from all of the applicable values. Once the calculations are completed, the individual results are compared against the mean \pm 2 standard deviations; if the result is out of range, the result is flagged for identification.

MA – Macroscopic Findings

The macroscopic findings domain is comprised of records related to a gross observation of specimens, a visual examination of individual tissues and organs, typically by the eye of the pathologist. Each record represents an individual finding per specimen and subject. It is possible that there are multiple findings for one specimen in a subject. Findings are classified into two categories – either the finding is unremarkable, and titled as such, or the “positive” finding, which has a test description of what was found. The term “normal” is not utilized as the reviewing pathologist determines if the positive finding is considered normal for the species or if there should be a remark for the examined specimen.

STUDYID	DOMAIN	USUBJID	MASEQ	MATESTCD	MATEST	MAORRES	MASTRESC	MASPEC	MALAT	MADTC	MADY
Participant-4	MA	Participant-4-0101	1	GROSPATH	Gross Pathological Examination	NORMAL	UNREMARKABLE	HEART		2020-06-08T09:36:16	29
Participant-4	MA	Participant-4-0101	2	GROSPATH	Gross Pathological Examination	Cyst(s), Left, Clear, 2-9mm, 1 (one)	Cyst(s)	KIDNEY	BILATERAL	2020-06-08T09:36:16	29
Participant-4	MA	Participant-4-0101	3	GROSPATH	Gross Pathological Examination	NORMAL	UNREMARKABLE	LARYNX		2020-06-08T09:36:16	29
Participant-4	MA	Participant-4-0101	4	GROSPATH	Gross Pathological Examination	NORMAL	UNREMARKABLE	LIVER		2020-06-08T09:36:16	29
Participant-4	MA	Participant-4-0101	5	GROSPATH	Gross Pathological Examination	NORMAL	UNREMARKABLE	LUNG/BRONCHUS	BILATERAL	2020-06-08T09:36:16	29
Participant-4	MA	Participant-4-0101	6	GROSPATH	Gross Pathological Examination	NORMAL	UNREMARKABLE	LYMPH NODE, TRACHEOBRONCHIAL		2020-06-08T09:36:16	29
Participant-4	MA	Participant-4-0101	7	GROSPATH	Gross Pathological Examination	NORMAL	UNREMARKABLE	SPLEEN		2020-06-08T09:36:16	29
Participant-4	MA	Participant-4-0101	8	GROSPATH	Gross Pathological Examination	NORMAL	UNREMARKABLE	TRACHEA		2020-06-08T09:36:16	29

Table 4.6: Example of MA domain with a remarkable finding for one specimen

In Table 4.6, subject number Participant-4-0101 is displayed with the specimens required at the time of necropsy. All specimens were “unremarkable”, with no findings, except the kidney which was found to have a cyst. The cyst was described with the modifiers of “left” regarding the left kidney, “clear” in color, “2-9mm” in size, and was quantified that there was a single cyst. In the MASTRESC column, it can be observed that the base finding, without the modifiers, was populated as “Cyst(s)”. If another subject has a cyst, multiple modifiers can be added, but all the

findings will be qualified as “Cyst(s)” rather than one remarkable finding per finding with all the modifiers.

Unremarkable vs. Remarkable Findings

Some findings are considered “normal” for a species or incidental by the type of treatment, like swelling and redness at the injection site. As such, a visualization was made to depict the presence of unremarkable findings, versus the number of positive findings, categorized by the standardized result (MASTRESC, Standardized Result in Character Format). If there is a high incidence of a certain type of finding in a specimen, this may indicate a result of the test article, be normal for the species, or be a factor of the environment. The visualizations can be passed along to the pathologist and study personnel to investigate the cause.

MI – Microscopic Findings

The microscopic findings domain is built of records related to the histopathology examination of specimens, a microscopic examination of individual tissues and organs, or parts of the tissue or organ. Each record includes a finding per specimen per subject; it is possible to have multiple findings for each specimen. Findings can either be unremarkable or considered a positive finding with a detailed description of the observation.

STUDYID	DOMAIN	USUBJID	MISEQ	MITESTCD	MITEST	MIORRES	MISTRESC	MIRESCAT	MICHRON	MIDISTR	MISPEC	MISEV	MIDTC	MIDY
1234567	MI	1234567-R0001	1	GHISTXQL	General Histopathologic Exam, Qual	Agonal congestion/hemorrhage, Present	CONGESTION/HEMORRHAGE	NON-NEOPLASTIC			LIVER		2020-05-28T10:20:32	22
1234567	MI	1234567-R0001	2	GHISTXQL	General Histopathologic Exam, Qual	NORMAL	UNREMARKABLE				KIDNEY		2020-05-28T10:20:32	22
1234567	MI	1234567-R0001	3	GHISTXQL	General Histopathologic Exam, Qual	NORMAL	UNREMARKABLE				LYMPH NODE, MESENTERIC		2020-05-28T10:20:32	22
1234567	MI	1234567-R0001	4	GHISTXQL	General Histopathologic Exam, Qual	NORMAL	UNREMARKABLE				OVARY		2020-05-28T10:20:32	22
1234567	MI	1234567-R0001	5	GHISTXQL	General Histopathologic Exam, Qual	Inflammation, alveolus/interstitium, minimal	INFLAMMATION	NON-NEOPLASTIC			LUNG	1 OF 5	2020-05-28T10:20:32	22
1234567	MI	1234567-R0001	6	GHISTXQL	General Histopathologic Exam, Qual	Infiltrate, inflammatory cell, minimal	INFILTRATE	NON-NEOPLASTIC			LIVER	1 OF 5	2020-05-28T10:20:32	22

Table 4.7: Example of MI domain with positive findings and unremarkable specimens

Table 4.7 displays some records for subject 1234567-R0001, which includes six specimens taken at necropsy. The MIORRES column displays the finding (or lack thereof), and MISTRESC is assigned according to the base finding that is in MIORRES, based on controlled terminology. Modifiers are not included in MISTRESC unless they are necessary for the

controlled terminology. “Minimal” in the MIORRES column is noted in the severity column (MISEV), which is also controlled terminology for a scale of #/3, #/4, or #/5. For this study, minimal is considered 1 out of 5 in the severity.

Unremarkable vs. Remarkable Findings

Some findings are considered “normal” for a species, or incidental by the type of treatment, like swelling and redness at the injection site. As such, a visualization was made to depict the presence of unremarkable findings, versus the number of positive findings, categorized by the standardized result (MISTRESC, Standardized Result in Character Format). In the MI domain, MISTRESC is subject to controlled terminology, which limits the number of findings that would be present. A high incidence of a certain finding may indicate toxicity, or may be a factor of the environment, a normal feature of the species, or another impact of the test article. The exact cause of the high incidence can be reviewed by the pathologist and study personnel for inclusion into the final study report.

Test Data

To test the program, two main types of data were tested: readily available data from a fit-for-use pilot, as well as a test dataset created with this intent in mind.

DART Fit-For-Use Pilot

As part of the development of the SEND standard, a fit-for-use pilot is typically run using 5-8 industry created studies and datasets for the proof of concepts to be released, and to prepare the FDA to review the datasets. The anonymized datasets and feedback are posted to the CDISC Wiki available to the public (Kramer, 2022).

The most recent fit-for-use pilot experiment was performed using the DART (Developmental and Reproductive Toxicology) v1.1 standard. Though DART is not the same as the SEND IG, it is a sub-development; many of the same parameters are available, including all the required parameters for the analysis program to run. SEND IG v3.1 datasets were not made publicly available at the time of the program's development, therefore the DART v1.1 datasets provided a suitable alternative. Six DART datasets are publicly available and were utilized to test portions of the program. Limitations of these datasets is that not all the datasets contained each domain to be tested, and that DART datasets only contain female animals. The datasets, however, do represent a real study conducted, and give an example of what significant findings may be noted.

TEST1 Dataset

To test the functionality of each of the domains together, a test dataset referred to by the study identification of "TEST1" was created with the knowledge of the requirements. A study was created that shows four dose groups (control, low, mid, high), with both sexes, and with a set of participants in groups 1 and 4 receiving a recovery, non-dose period.

The bodyweight data, for example, was created by utilizing a range of typical body weights for the species, separated slightly by sex. The animals within the study were randomized to be within the given range so that they all included a body weight within the "normal" range at baseline. For each additional observation, the last value was taken, and a random number generator was used to add or subtract up to 10% of the body weight for each period of seven days. At the last observation of the study, the randomized number generator was increased to up to a 13% difference from the previous observation. This artificially created a dataset that would

be within range, but each subject was individually “gaining” or “losing” weight at the subject of a random number generator to eliminate any human bias.

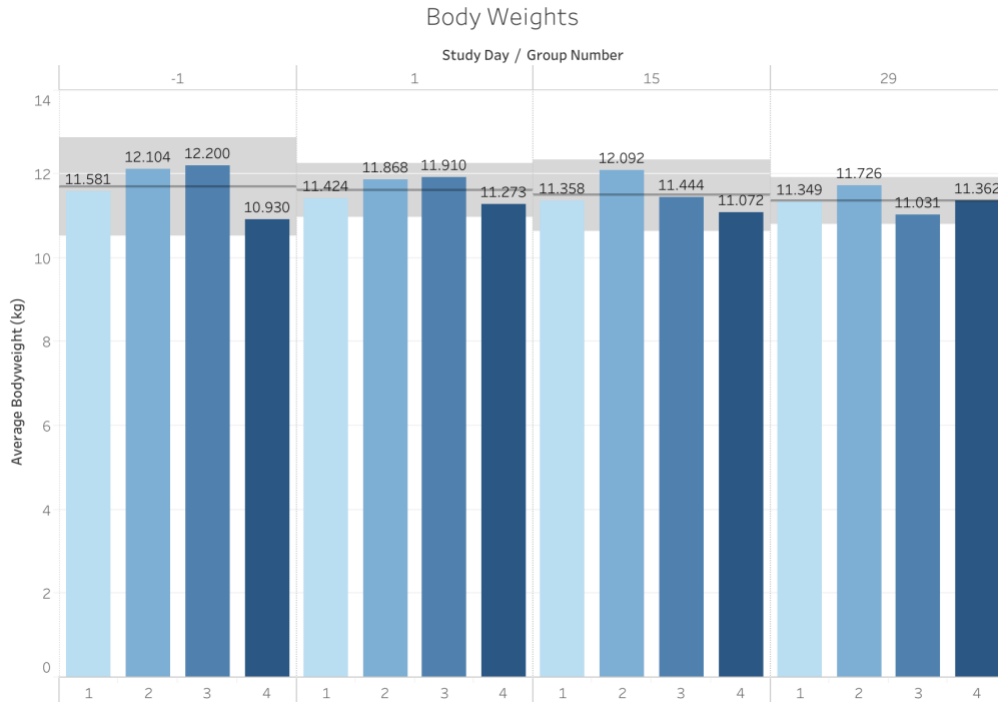


Figure 4.1: Body weight means of TEST1 study by a group with 90-110% of control group weight highlighted

Figure 4.1 visualizes the body weight data that was included in the TEST1 dataset. Study days -1, 1, 15, and 29 were included. The mean of each group is displayed by the text label above it. The mean of group 1 (the control group) is used as a line across the day, and from 90-110% of the control group weight is highlighted. Group 4 on day -1 approaches the bottom of the acceptable range at close to less than 10% less than the mean of group 1. On days 15 and 29, group 2’s mean is approaching the upper level of acceptability based on the control group’s mean.

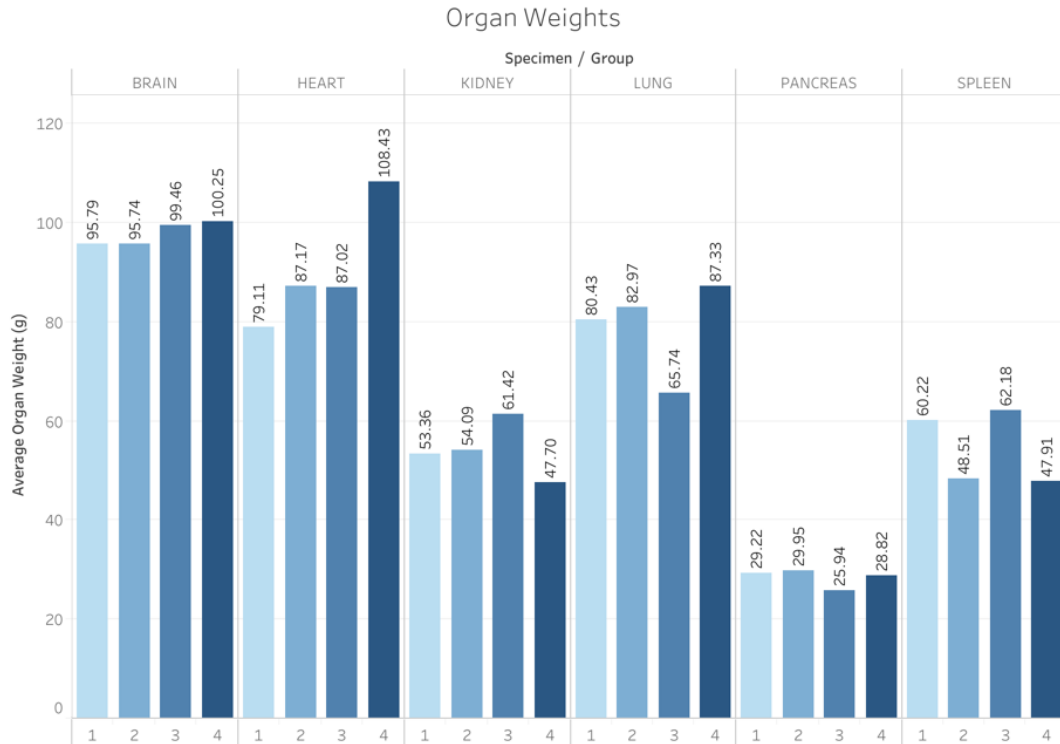


Figure 4.2: Selected organ weight means by group in the study TEST1

Select tissues (organs) were included in figure 4.2 for the groups 1-4 to summarize the mean weight per organ. The brain weight is fairly similar across the groups. The heart and lungs, however, have fairly different means in certain groups. Group 4 has a heart weight that is up to 30 grams more than the control group; group 3 has a mean lung weight that is 15 grams less than the control group's mean lung weight.

Analysis and Discussion

A program was successfully created using the R and Python languages, first to solidify a standard of input and then to analyze the data that is presented. When submitting a SEND dataset to the analysis program, the program will only detect the .xpt file format, and accept domains with the abbreviations BW, LB, OM, MA, or MI. These files are individually analyzed using the accompanying DM and TX domains for context for group (dose level), sex, and if the group was a designated control group for the study. Any findings that were considered out of range by a range of analysis types are recorded and sent to further evaluation and grouping of the findings by subject number and domain.

After analysis, a two-part deliverable was created, a file called “report.txt” that describes the portions of analysis that were not able to be completed, and a file called “SignificantFindings.xls”, referred to as the “significant findings file” hereafter. The significant findings file contains identifying information on each of the outlying values found by the program, as well as the visualizations for the MA and MI domains. Each finding includes at least the domain, animal number, specimen or day, and the type of calculation that was utilized to identify it as a finding. When the finding is identified multiple times, such as animal 1008 body weight is found to be out of range on day 22, there will be multiple findings included in one line of findings, separated by comma in regard to what percentage or standard deviation it falls outside of the acceptable range, and the calculation type utilized, also separated by comma.

USUBJID	Domain	Identifier	Value	Method
TEST1-1001	BW	BWDY -1	-10.878	Percent Change from Control (%)
		BWDY -8	10.011	Percent Change from Baseline (%)
		BWDY 1	-14.346	Percent Change from Control (%)
		BWDY 15	-14.118	Percent Change from Control (%)
		BWDY 22	-18.428, -11.489	Percent Change from Control (%), Percent Change from Baseline (%)
		BWDY 29	-14.426	Percent Change from Control (%)
		BWDY 8	-19.04, -11.08	Percent Change from Control (%), Percent Change from Baseline (%)
TEST1-1002	BW	BWDY -8	10.834	Percent Change from Control (%)
		BWDY 29	15.452	Percent Change from Control (%)
TEST1-1003	BW	BWDY 1	13.555	Percent Change from Control (%)
		BWDY 15	19.722, 11.495	Percent Change from Control (%), Percent Change from Baseline (%)
		BWDY 22	27.867, 18.854	Percent Change from Control (%), Percent Change from Baseline (%)
		BWDY 29	19.334, 13.03	Percent Change from Control (%), Percent Change from Baseline (%)
		BWDY 8	23.828, 16.505	Percent Change from Control (%), Percent Change from Baseline (%)

Table 5.1: Bodyweight findings for three subjects in the BW domain with methods listed

Table 5.1 is an example of the significant findings document where body weights were identified as being out of range. The subject is identified in the first column, and the body weight domain is represented in the second column. All findings for the subject were grouped together for analysis of a single subject. The bodyweight for subject TEST-1001 on day 22 was identified as being out of range by two different detection methods, both by a change from the control group, and by the percent the body weight changed from the baseline observation (more than 10%). Other findings were identified by one method or the other, either from the change from the control, or the change from the baseline for the subject in question.

USUBJID	Domain	Identifier	Value	Method
TEST1-4003	BW	BWDY -8	-11.736	Percent Change from Control (%)
		BWDY 22	16.794, 14.3	Percent Change from Control (%), Percent Change from Baseline (%)
		BWDY 29	23.442, 23.101	Percent Change from Control (%), Percent Change from Baseline (%)
	OM	BRAIN	>=2	Standard Deviation range by Control group
		GLAND, ADRENAL	>=2	Standard Deviation range by Control group
		HEART	>=2	Standard Deviation range by Control group
		KIDNEY	>=2	Standard Deviation range by Control group
		LIVER	>=2	Standard Deviation range by Control group
		LUNG	>=2	Standard Deviation range by Control group

Table 5.2: Single subject findings for the BW and OM domain

Subject TEST1-4003 is shown in table 5.2 with both BW and OM domain findings. The body weight was identified as being below an acceptable limit for both days 22 and 29 by both the baseline and change from control methods. Additionally, six of the organs listed in the OM

domain for the subject were identified as being “out of range” as the weight fell outside of the mean \pm 2 standard deviations of the control group.

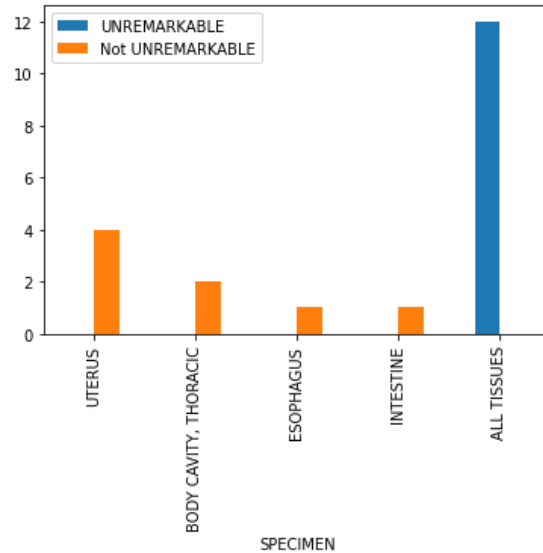


Figure 5.1: Remarkable and unremarkable findings in a group of animals; specimens were only noted individually if there were findings

Figure 5.1 is a visual description of the number of findings for a group of subjects in their respective organs. When there were no remarkable findings for the subject, the “ALL TISSUES” specimen was populated, and “UNREMARKABLE” was populated in MASTRESC. When there were other findings, the findings were quantified in the visualization per specimen. In the group this was calculated for, there were 12 records that were “UNREMARKABLE” that were always populated with the specimen of “ALL TISSUES”; there were four findings (in orange, “Not UNREMARKABLE”) for the uterus, two findings for the thoracic body cavity, one finding for the esophagus, and one finding for the specimen of “INTESTINE”. The positive findings in the uterus may be examined more closely as there were 16 subjects in the group, and 25% of the group had findings in the uterus.

The program has, in running only .xpt files, identified outlying data points in under 1 second per study. From a timing aspect, the program can run on five domains before a human could open each of the five files, increasing efficiency for the person responsible for interpreting the data. The efficiency increases exponentially with the number of records a study contains – as studies can contain hundreds of thousands of records, even small amounts of identification and analysis techniques can save hours of manual reviewing and provide a starting point for the approval of the study results.

The program increases readability for the responsible personnel if they are not familiar with the concept of SEND, which is new to the industry, and specialized mostly in the areas of reports. The significant findings file is produced in a way that it can be read by someone familiar with how non-clinical datasets are produced, as it calls out individual identifying information. The significant findings file can be read across a line as “[animal number] is out of range for [test] on [day] by [percent/standard deviation] by [calculation]”.

The program is currently only designed to handle SEND 3.1 based datasets, which limits the scope of datasets that can be evaluated. Of the datasets, only five domains are currently integrated, though, with more available datasets and exploration time, other domains, such as food consumption and respiratory test results can be included.

Conclusions

This study has provided a tool for data analysis in pharmaceutical studies that can be arranged in a standardized format with the intention of detecting toxicity early for the improved safety and efficacy of animals. Toxicity in animals is characterized by a number of changes, some of which can be identified by observations measured and presented in the SEND dataset either in-life or postmortem. If these endpoints are identified in-life, undue distress can be prevented in the animals. If the endpoints are identified postmortem, further study protocols can be amended if necessary to prevent the same outcome. With a tool for identification, the analysis of these endpoints is accelerated, and the results can be presented in a more user-friendly fashion.

The analysis program currently evaluates the endpoints of the most common domains in preclinical studies; additional domains that involve findings, such as clinical signs, food and water consumption, and safety pharmacology domains like respiratory test results and vital signs, can be implemented in similar methods of review for a full study analysis of outliers. Large amounts of outlying data may indicate a reason for discontinuance of future studies in the interest of safety and looking to other developments with the conclusions of the study with outliers.

The program is currently limited to the five domains listed in the methodology, which are not the only domains used for determining toxicity, however, they do provide insight into the subject's reaction to the test articles. Historical control data was not incorporated as there is industry standard of the control data; every research lab has its own host of historical data that

they consider when evaluating for toxicity. Future work may include the ability to calculate outliers from this data, but some LIM systems already do include this functionality.

One functionality that would be beneficial to the study, but require the opinions of multiple pathologists, would be to incorporate a set of standards outside of the statistical analysis for flagging toxicity, i.e., a way to convert the MI domain findings to controlled terminology, and have a list of terminologies that are considered a sign of toxicity per species. There are some categorizations of findings that are considered normal for one species but may indicate a change from the test article in another. The limitation of this possible research is that all findings are extremely study dependent – it may be normal for a finding of a tumor within a subject. The subject was first implanted with tumor cells, but the test article is a tumor inhibitor, meant to reduce the number of tumor cells in the body.

As the SEND standard develops, it is possible that this program or the research performed can evolve and begin to evaluate other domains, as well as evaluate other standards, like the sub-standards of SEND involving developmental and reproductive toxicology (SEND-DART v1.1 accepted per March 2023). As SEND is developed off of the SDTM standard, the tool can be revised in variable naming, and specific percent changes to relate to clinical trials and provide an overview of changes in laboratory tests as the trials are ongoing.

The program created will hopefully be used in the future to perform interim study monitoring, checking how each record has changed across not only the subject, but the group for the entire study. By monitoring the overall changes, the performing laboratory can be ready at the end of the study to perform the lab-specified statistical analysis to prove if the changes were

due to an impact of the test article, or if the changes were within limits and were up to another factor such as the environment or another biological process within the species' body.

Research studies that are already presented in an electronic exchange format can be analyzed for anomalous data without the context of the study. Formal recommendations must include the background and a person that is an expert in the field, but identifying the data points has become a one step process of opening the program and inserting a study number, as opposed to doing manual calculations using the limitations of the human's interpretation.

References

Hacking the Thesis. (n.d.). The Ohio State University. Retrieved February 16, 2022, from <https://u.osu.edu/hackingthethesis/managing-stuff/your-content/outline/>

SAS Transport File Format (XPORT) Family. Retrieved May 8, 2023, from <https://www.loc.gov/preservation/digital/formats/fdd/fdd000464.shtml>.

Ballentine, C. (1981, June). Taste of raspberries, taste of death: the 1937 elixir sulfanilamide incident. *FDA Consumer*. Retrieved from <https://www.fda.gov/files/about%20fda/published/The-Sulfanilamide-Disaster.pdf>

Center for Drug Evaluation and Research, & Kux, L., Electronic Study Data Submission; Data Standards; Support for Standard for Exchange of Nonclinical Data Implementation Guide Version 3.1. Federal Register. Retrieved from <https://www.federalregister.gov/documents/2017/08/21/2017-17567/electronic-study-data-submission-data-standards-support-for-standard-for-exchange-of-nonclinical>

Center for Drug Evaluation and Research. (2014, October 23). *Types of Applications*. U.S. Food and Drug Administration. <https://www.fda.gov/drugs/how-drugs-are-developed-and-approved/types-applications>

Center for Drug Evaluation and Research. (2022a, June 21). *Center for Drug Evaluation and research: CDER*. U.S. Food and Drug Administration. <https://www.fda.gov/about-fda/fda-organization/center-drug-evaluation-and-research-cder>

Center for Drug Evaluation and Research. (2022b, July 20). *Investigational New Drug (IND) Application*. U.S. Food and Drug Administration. <https://www.fda.gov/drugs/types-applications/investigational-new-drug-ind-application>

Charles River. (n.d.). SEND: Standard for Exchange of Nonclinical Data. Retrieved from https://www.criver.com/sites/default/files/resources/doc_a/StandardforExchangeofNonclinicalDataSENDe-Guide.pdf

Commissioner, O. of the. (2018, January 4). *Step 2: Preclinical research*. U.S. Food and Drug Administration. <https://www.fda.gov/patients/drug-development-process/step-2-preclinical-research>

Commissioner, O. of the. (2022, January 18). *What does FDA regulate?*. U.S. Food and Drug Administration. <https://www.fda.gov/about-fda/fda-basics/what-does-fda-regulate>

Commissioner, O. of the. (2023, January 30). *FDA history milestones*. U.S. Food and Drug Administration. <https://www.fda.gov/about-fda/fda-history/milestones-us-food-and-drug-law>

Environment Directorate. (2000, November). Guidance document on the recognition, assessment, and use of clinical signs as humane endpoints for experimental animals used in safety evaluation. Paris; ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. Retrieved from https://ntp.niehs.nih.gov/iccvam/suppdocs/feddocs/oced/oced_gd19.pdf

FDA. (2015, April 7). *Definitions*. U.S. Food and Drug Administration. <https://www.fda.gov/inspections-compliance-enforcement-and-criminal-investigations/fda-bioresearch-monitoring-information/definitions>

Good laboratory practice (GLP) - OECD. Organisation for Economic Co-operation and Development. (n.d.). <https://www.oecd.org/chemicalsafety/testing/good-laboratory-practiceglp.htm>

Hajar, R. (2011). Animal Testing and Medicine. *Heart Views*, 12(1), 42. <https://doi.org/10.4103/1995-705x.81548>

Hughes, J., Rees, S., Kalindjian, S., & Philpott, K. (2011). Principles of early drug discovery. *British Journal of Pharmacology*, 162(6), 1239–1249. <https://doi.org/10.1111/j.1476-5381.2010.01127.x>

Investigators, sponsors, and sponsor-investigators. Investigators, Sponsors, and Sponsor-Investigators | Social Science Research Institute. (n.d.). <https://ssri.psu.edu/investigators-sponsors-and-sponsor-investigators>

Kramer, L. A. (2022, May 17). SENDIG-DART v1.1 Fit For Use Pilot. Retrieved from <https://wiki.cdisc.org/display/NSFFUW/SENDIG-DART+v1.1+Fit+For+Use+Pilot?desktop=true&oname=view-file>

Meadows, M. (2006). Promoting safe and effective drugs for 100 years. *FDA Consumer*. Retrieved from <https://www.fda.gov/files/Promoting-Safe-and-Effective-Drugs-for-100-Years-%28download%29.pdf>

Pandas. (n.d.). Version (1.5.3). *pandas*. Retrieved from <https://pandas.pydata.org>.

Send. CDISC. (n.d.). <https://www.cdisc.org/standards/foundational/send>

Sun, D., Gao, W., Hu, H., & Zhou, S. (2022). Why 90% of clinical drug development fails and how to improve it? *Acta Pharmaceutica Sinica B*, 12(7), 3049–3062. <https://doi.org/10.1016/j.apsb.2022.02.002>

Van Norman, G. A. (2016). Drugs, devices, and the FDA: Part 1. *JACC: Basic to Translational Science*, 1(3), 170–179. <https://doi.org/10.1016/j.jacbts.2016.03.002>

Vargesson, N. (2015). Thalidomide-induced teratogenesis: History and mechanisms. *Birth Defects Research Part C: Embryo Today: Reviews*, 105(2), 140–156.
<https://doi.org/10.1002/bdrc.21096>

White Junod, S. (2008). FDA and Clinical Drug Trials: A Short History. *A Quick Guide to Clinical Trials*. Retrieved May 8, 2023, from
<https://www.fda.gov/media/110437/download>.

Wood, F., & Kramer, L. A. (2011). The Standard for the Exchange of Nonclinical Data (SEND): History and Basics. *PharmaSUG 2011*. Retrieved from
<https://www.lexjansen.com/pharmasug/2011/CD/PharmaSUG-2011-CD14.pdf>

Appendix A – Definitions

Some terms are used frequently, either in name or abbreviation, that are common in the pharmaceutical data analysis industry; expanded names and explanations are defined below as they relate to this study.

1. Center for Drug Evaluation and Research (CDER) – Department of FDA that regulates over the counter and prescription drugs
2. Contract Research Organization (CRO) – Laboratory that is contracted by the sponsoring organization to perform the physical research of a study
3. Food and Drug Administration (FDA) – U.S. government agency responsible for regulating food, drugs, biologics, etc.
4. Good Laboratory Practice (GLP) – Set of regulations to ensure quality and integrity of non-clinical safety studies
5. Investigational New Drug application (IND) – Preclinical data to permit an assessment as to whether the product is reasonably safe for initial testing in humans
6. Laboratory Information Management System (LIMS) – Program or data recording tool to note individual data samples, including records of dosing, clinical sign findings, and laboratory test result measurements
7. Non-clinical – In vivo study in which a test article is studied in a test system, excluding utilizing human subjects, synonymous with preclinical in this study
8. Organisation for Economic Co-operation and Development (OECD) – Compliance monitoring organization, type of good laboratory practice
9. Preclinical – In vivo research performed on animals to provide detailed information on dosing and toxicity levels, synonymous with non-clinical in this study

10. Standard for Exchange of Nonclinical Data (SEND) – collection and presentation of nonclinical data in a consistent format
11. SEND Implementation Guide (SENDIG) – document providing implementation guidance for the SEND standard, including rules, recommendations, and examples
12. Sponsor – Individual, pharmaceutical company, government agency, academic institution, private organization, or other organization that is responsible for the initiation of a clinical investigation. A sponsor is the owner of the test article.
13. Test Article – Substance or mixture administered in a test system in the study for a research or marketing application