Paper SA10

Regulatory Submissions – Exciting, Challenging or Both

Ganeshchandra Gupta, Ephicacy Consulting Group, Inc., Piscataway, NJ, USA
Randi McFarland, Ephicacy Consulting Group, Inc., Rocklin, CA, USA

ABSTRACT

It's exciting to help submit the final package going to regulatory authorities in hopes of product approval! However, challenges always arise preparing regulatory submissions from the data collection and cleaning to data integration and analyses with so many stakeholders involved. What kind of strategies and best practices help us navigate these challenges to reduce potential issues and delays? By identifying potential risks and addressing them early, we can streamline the submission process for data packages. This presentation will share best practices and strategies for programmers to overcome challenges for providing the required elements for a successful submission.

INTRODUCTION

Pharmaceutical companies run clinical trials in hopes of achieving regulatory approval of their products for widespread medical use. Regulatory approval depends not only on the sponsor's clinical trial development, implementation and data capture oversight, but also on the quality of the data processing and analysis. Statistical programming is a fundamental part of quality data submissions. Well-planned coordination with data management increases the likelihood of accurate results while minimizing the processing timeline. Statisticians provide the plans for the analyses and early collaboration provides the programmers with more understanding of the data analysis requirements. Challenges may arise from complexities in the clinical trials' data, the data integration, and the need to adhere to strict regulatory guidelines. By identifying potential challenges, strategies and best practices can be implemented to mitigate those issues and streamline the submission process.

WORKING WITH DATA MANAGEMENT

To produce the analyses and supporting documentation for a regulatory submission, planning should be initiated between the data management and analysis teams. Statistical programmers work with statisticians to understand the planned analyses and study integration requirements. They are expected to provide accurate, reliable analyses that follow the analysis plan, industry data standards (CDISC) and regulatory (FDA) guidance documents. To develop statistical analyses, programmers rely on Data Managers to follow their standard data review and cleaning processes to produce accurate data within the planned timelines. Data Managers ensure that appropriate visits per patient and data (CRF pages) per visit are available and query clinical sites regarding questionable and missing data. While some data may remain missing, other data may be present but not acceptable for analysis. It can be challenging to ensure important data is analyzable, as data management may not always view the data from the programmer's perspective. Statistical programmers must clearly communicate to data managers the requirements for receiving the data as expected.

For example, visit data and the corresponding dates must match chronologically to assess the treatment impact on blood pressure or pain scale ratings or other parameters.

If we assume patient 201 has the following data, he shows no improvement if we sort by visit and compare the first and last pain ratings. However, if we re-sort the data by date, improvement from score of 10 down to 3 is shown.

TABLE 1. TABLE OF PATIENT 201 WITH INVALID DATES FOR VISITS

Visit	Collection Date	Pain Rating
1	2022-02-22	10
2	2022-07-22	3
3	2022-04-23	8
4	2022-05-24	7
5	2022-03-23	10

This is a simple example where the data manager has confirmed all data has been provided, however, they may not realize the potential impact on the analysis if dates and visits are not in sync. Explaining why it is important for visits and dates to

match chronologically will help data managers understand the purpose of the analysis and find these or similar issues in other datasets.

Here is another example of problematic data.

TABLE 2. TABLE OF ADVERSE EVENTS WITH INVALID DATES

Subject	Adverse Event (AETERM)	Event Start Date	Event End Date
30010-001	Headache	30AUG2023	17JUN2023
30501-003	Nausea	05NOV2022	04NOV2022
20221-005	Hypertension	10FEB2023	01JAN2023
20252-002	COVID-19	16MAY2024	13MAY2024

In this scenario, the AE dataset contains these adverse events where the event end date precedes the start date. This is a data integrity issue that needs addressing. Accurate date sequencing is fundamental for ensuring the validity of our data. When the end date is before the start date, it indicates a possible error in data entry or processing. For computations in statistical analyses and reporting, such as calculating event duration, survival rates, or other key metrics, the dates must align correctly. Incorrect date sequences can lead to inaccurate results, affecting the reliability of the analysis and conclusions. Data managers should be encouraged to review the data with these considerations in mind.

Strategies working with the Data Management (DM) team involve frequent project communication, feedback for ongoing studies and improving their understanding of data analysis requirements. This collaboration starts early in the project and continues regularly throughout, planning data cuts, and data reviews enabling the prompt identification and resolution of data discrepancies. Performing early data reviews allows statistical programmers to detect data issues like missing or inconsistent data values, ensuring the data's robustness for the analyses. These systematic data reviews also help the programmers and statisticians to verify the accuracy of data transformations and derivations. This programming and data management collaboration ensures the quality of the data needed to mitigate submission delays. Data management’s understanding of data analysis requirements and providing early data cuts are necessary to streamline the process.

INTEGRATION PLANNING AND DEVELOPMENT

Integration of data from various studies or sources is another critical aspect managed by statistical programmers. Data integration and harmonization needs to be carefully planned and executed to perform the data analyses specified in the integrated Statistical Analysis Plan (iSAP). When pooling data, complicated situations must be considered such as differences in data collection, data coding and dictionaries or data standards used. Planning which data is to be analyzed and identifying the pooling, re-coding and up-versioning of that data are critical challenges to be addressed ensuring successful harmonization.

First, data integration requires a detailed integrated Statistical Analysis Plan (iSAP) that outlines the intended analysis strategies and data pooling methods. Statistical programmers rely on the iSAP, mock tables and data to create analysis dataset specifications, integrated analysis datasets and summary tables. The programmer works with the statistician to understand the analyses and integration requirements. It's crucial to clearly define primary endpoints like Overall Survival (OS) and Progression-Free Survival (PFS), and to specify the statistical models that will be employed for these measures. If the iSAP and mocks do not provide enough detail to combine the data appropriately and produce required analysis variables, the specifications and datasets will need to be corrected. The summary tables must be clearly defined to prevent incorrect analyses leading to questions, review and rework, potentially postponing the draft analysis review. If data is not provided at planned intervals, development of analysis programs may be delayed. In both cases, communication and coordination with data managers and statisticians are critical.

Some submissions involve generating hundreds of tables, many of which are repeated. One strategy is to organize the mocks using internal identification numbers that signify the types of analyses where the standard mock is used.

For example, a typical mock-up for an efficacy table might be labeled as “TE-01,” while a corresponding figure that visually represents the same data would be designated as “FE-01.” Safety summary tables are generally marked as “TS-01,” “TS-02,” and so forth. To distinguish between Integrated Summary of Safety (ISS) and Integrated Summary of Efficacy (ISE) tables, we can add an SS or SE prefix. In the case of an ISE, “SETE-01” and “SEFE-01” denote the efficacy table and figure, respectively. For ISS safety tables and figures, we follow the same structure, using labels such as “SSTS-xx” for tables and “SSFS-xx” for figures.

TABLE 3. TABLE OF MOCK ASSIGNMENTS FOR ISS AND ISE

Program Name	ICH Table Number	Mock Table Number	Include in ISS	Include in ISE
tdemsum	14.1.1	SSTS-01	X	X*
tcharsum	14.1.2	SSTS-02	X	X*
taeall	14.3.1.1	SSTS-03	X	X*
taesoc	14.3.1.2	SSTS-04	X	
tlabhem	14.3.2.1	SSTS-07	X	
flabvis	14.3.2.2	SSFS-04	X	X*
teff01	14.2.1	SETE-01		X
feff01	14.2.1.1	SEFE-01		X

** The same mock from the ISS is being used for the ISE in this case*

A table could show the assignments of these standard mocks to assign to each ICH numbered table such as 14.1.1, 14.2.1, 14.3.1, etc. Some standard ISS table mocks may be used in the ISE or Risk Management Plan (RMP) portions of the submission, providing clarity to the statistical programmer, and may reduce the number of mocks and analysis programs required.

With a clear SAP and mock tables in place, the statistical programmer can develop detailed pooled dataset specifications ("specs"). These specs outline the data pooling methods, including new variables for grouping or re-coding. By creating these programming specifications, programmers ensure consistency and accuracy across integrated datasets. However, integrating from different studies often presents challenges as studies may collect varying data and use different algorithms, such as baseline definitions or treatment lag times (length of time the treatment is considered active). To address this, creating a table of these differences by study or analysis (ISS, ISE, etc.) can help identify where algorithms should be updated to a standard method or highlight known differences.

TABLE 4. TABLE OF ALGORITHM DIFFERENCES ACROSS STUDIES AND SUBMISSIONS

Algorithms	Data	Study 1	Study 2	Study 3	ISS
Baseline	Vital sign measurements	Last value prior to visit 1	Day 1 value	Last visit of run-in period	Last visit prior to Day 1
Baseline	ECG measurements	Average of last Triplicate prior to visit 1	Day 1 value	Last visit of run-in period	Last visit prior to Day 1
Treatment Lag time	On-treatment Adverse Events	7 days post last dose	14 days post last dose	10 days post last dose	7 days post last dose

There may also be differences in codelists applied and in dictionaries or CDISC standard versions used. If different versions of MedDRA dictionaries, or SDTM/ADaM standards are used, up-versioning to the latest dictionary or SDTM/ADaM standard may be required. This ensures consistent pooling of adverse events, data using different SDTM controlled terminology, or different variables capturing the same data. These differences can create challenges for the programmers during data integration.

Here is an example of verbatim terms that have different coded terms based on the MedDRA dictionary version.

TABLE 5. TABLE OF MEDDRA CODING DIFFERENCES BY DICTIONARY VERSION

Adverse Event (AETERM)	Preferred Term (MedDRA v24.1)	Preferred Term (MedDRA v27.0)
Eyelid Papilloma	Blepharal Papilloma	Papilloma
Left Knee Pain	Osteoarthritis	Arthralgia
Malaise With Vaccination	Vaccination Complication	Immunisation Reaction
Tachycardia Post Covid-19 Infection	Post-acute Covid-19 Syndrome	COVID-19

In planning and developing data integration, differences in algorithms, coding, standards and dictionaries used across studies must be documented and possibly updated for consistency. This facilitates accurate analyses using different data sources and provides traceability to the original sources. Statisticians review integrated datasets to ensure data pooling, re-coding, new variables, and algorithms are accurate for the analyses.

DOCUMENTATION

When statistical programmers develop regulatory submission packages, they must ensure compliance with regulatory submission guidance documents and provide the comprehensive statistical programming documentation expected. The guidance documents from the FDA[1,2,3] and standards and documentation templates from CDISC[4] and PHUSE[5] organizations are available for reference. The Technical Conformance Guide (TCG)[1] is a document produced by the FDA that specifies data submission requirements. These requirements include following the CDISC SDTM and ADaM Implementation Guides (IGs) and the Study Data Standardization Plan (SDSP) to create study data and plan submissions. Individual study data documentation is based on the PHUSE cSDRG and ADRG templates, with the inclusion of the CDISC Define.xml. While these documents are mandatory for the FDA, other regulatory agencies may have their own guidance documents. The Japanese regulatory authority, PMDA, for example, has their own Technical Conformance Guide[6]. Without expertise in the use of these documents, and experienced programmers to implement them, challenges may arise. Incomplete documentation and data, or unmet minimum standard requirements can potentially hinder a successful submission.

Providing thorough and well-organized documentation is essential for regulatory submissions. Statistical programmers are responsible for documenting programming codes, analysis and integrated data specifications, and analysis methods carefully. Data reviewer's guides describe the submitted data in detail. This documentation provides clarity, transparency and reproducibility, making it a critical component of regulatory submissions and FDA audits. By documenting every step of the analysis process, programmers contribute to the reliability of the clinical data analyses.

BEST PRACTICES

A wide variety of best practices and effective methods are used to plan and develop regulatory submissions by statistics, programming and data management departments. Forming project teams of experienced data managers, statistical programmers, and statisticians enhances collaboration. These teams work closely, sharing insights and expertise to improve accuracy of the collected data and the efficiency of the data quality process. Team meetings should be held regularly to discuss and document new interpretations of the protocol or SAP, data requirements for new analyses and impactful changes of the data for future reference.

Statisticians provide clear guidance in the integrated statistical analysis plans (iSAPs) and mock tables for the submission. Programmers develop reusable macros to provide consistency across the analyses. Implementing rigorous validation procedures including edit checks by data managers, independent double programming by statistical programmers and peer review by statisticians will identify discrepancies and verify results. Expertise is required to develop analysis dataset specifications, which integrate multiple sources of data appropriately and clearly define how the datasets are to be created. Although the Pinnacle 21 (P21) tool is not designed to validate pooled data, it is often used to validate integrated datasets for compliance with CDISC standards. The industry is expected to follow the FDA or other regulatory guidance, CDISC standards, PHUSE SDTM and ADaM data reviewer's guides, and other standard templates. It is imperative to use these documents to produce the submission's analyses, data packages and documentation expected by regulatory agencies.

CONCLUSION

Clinical trials' data approval is a complex process that requires meticulous planning, robust strategies and seamless collaboration among data managers, statistical programmers and statisticians. By identifying risks and challenges and utilizing strategies and best practices we can mitigate risks and streamline the submission process. Planning for key components including early and frequent communication, thorough submission process planning, detailed documentation, and integrated project teams leads to success. These strategies and best practices not only ensure compliance with regulatory guidelines but enhance the reliability and validity of the clinical trial data and facilitate timely and successful regulatory submissions.

REFERENCES

[1] Study Data Technical Conformance Guide - Technical Specifications Document (Oct 2024) [Study Data Technical Conformance Guide_v5.9_Oct 2024]

[2] Data Standards Catalog (April 2024) [http://www.fda.gov/forindustry/datastandards/studydatastandards/default.htm]

[3] Providing Regulatory Submissions in Electronic Format — Standardized Study Data (June 2021) [https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM292334.pdf]

[4] Clinical Data Interchange Standards Consortium (CDISC): SDTM and ADaM Standards and Implementation Guides [https://www.cdisc.org/standards]

[5] PHUSE Submission templates including BIMO, Data reviewer's Guides and Study data Standardization Plan (SDSP) [https://advance.phuse.global/display/WEL/Deliverables]

[6] PMDA Technical Conformance Guide (April 2024) [https://www.pmda.go.jp/files/000267935.pdf]

CONTACT INFORMATION

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

Author Name: Ganeshchandra Gupta
Company: Ephicacy Consulting Group, Inc.
Address: 33 S Wood Ave Suite 600, Iselin, NJ 08830
Cell Phone: (848) 219-8131
Author Email: ganeshchandra.gupta@ephicacy.com
Website: https://www.ephicacy.com

Co-author Name: Randi McFarland
Company: Ephicacy Consulting Group, Inc.
Address: 33 S Wood Ave Suite 600, Iselin, NJ 08830
Cell Phone: (916) 580-4414
Co-author Email: randi.mcfarland@ephicacy.com
Website: https://www.ephicacy.com