**End-to-End SAS Programming for Regulatory Submissions and Clinical Trial Analysis**

**Executive Summary**

This project demonstrates the application of advanced SAS programming techniques to support late-stage clinical trials and regulatory submissions. The primary objectives included creating analysis-ready datasets, automating deliverables, and ensuring adherence to industry standards such as CDISC ADaM. Key contributions involved data cleaning and derivation, TLF generation, Kaplan-Meier survival analyses, and the preparation of regulatory submission documents. This work facilitated the efficient analysis of clinical data, improved process compliance, and enabled actionable insights for decision-making.

**Introduction**

**Background**

Effective statistical programming plays a crucial role in late-stage clinical development, driving high-quality analyses and deliverables for regulatory submissions. This project focuses on preparing and analyzing adverse event (AE) datasets to assess safety, stratify outcomes, and generate submission-ready outputs. The work aligns with the objectives of improving patient safety, optimizing clinical workflows, and adhering to regulatory standards.

**Scope**

The project includes:

1. Data cleaning, derivation, and standardization.
2. Automation of TLF (Tables, Listings, and Figures) generation.
3. Kaplan-Meier survival analysis.
4. Creation of regulatory submission documents, such as Define.XML metadata, XPT files, and comprehensive PDF reports.

**Methodology**

**1. Data Preparation**

* **Cleaning and Validation:**
  + Removed duplicate records and resolved missing values using SAS procedures (PROC SORT, PROC FREQ).
* **Variable Derivation:**
  + Derived key variables such as ADT (Analysis Date) and ASTDY (Analysis Start Day) to ensure analysis readiness.
* **Standardization:**
  + Mapped severity (AESEV) and relationship (AEREL) categories into numerical formats for consistency across analyses.

**2. Automation**

* Developed reusable macros to streamline repetitive tasks, including:
  + **Adverse Event Frequency Tables:** Macro-generated tables summarized AE frequencies by severity and body systems.
  + **Visualizations:** Automated creation of bar charts and Kaplan-Meier survival curves stratified by age, sex, and treatment arm.

**3. Survival Analysis**

* Conducted Kaplan-Meier analyses using PROC LIFETEST:
  + Stratified by demographics (e.g., age group, sex) to identify patterns in survival probabilities.
  + Extracted median survival times and evaluated trends for decision-making.

**4. Regulatory Submission**

* **Metadata Preparation:**
  + Extracted dataset metadata for Define.XML using PROC CONTENTS and exported it to Excel.
* **Dataset Export:**
  + Created transport files (XPT) using LIBNAME for FDA compliance.
* **Comprehensive Reporting:**
  + Generated a structured PDF submission report using ODS PDF, integrating key analyses, adverse event listings, and Pinnacle 21 validation summaries.

**Key Findings and Insights**

**Kaplan-Meier Survival Analysis**

* Survival probabilities were stratified by age groups, with the "Under 30" cohort demonstrating the highest survival rates. Older age groups ("50+") experienced steeper declines, highlighting age as a factor in clinical outcomes.
* Treatment arms revealed significant differences, with one arm maintaining consistently higher survival probabilities, indicating greater efficacy.

**Adverse Event Analysis**

* Frequency tables showed key safety concerns across treatment arms and demographic groups.
* Visualizations provided clear insights into the distribution of adverse events by severity and body systems.

**Strategic Impact**

**Efficiency and Automation**

* Automated workflows saved significant programming time and minimized errors, supporting scalable processes for future projects.

**Compliance and Quality**

* Adherence to CDISC ADaM standards ensured regulatory readiness and alignment with industry best practices.

**Real-World Applications**

* Findings informed clinical decision-making by providing demographic-specific insights and robust safety profiles.

**Conclusion**

This project demonstrates the integration of SAS programming expertise with strategic objectives in clinical development. The ability to automate processes, conduct survival analyses, and prepare regulatory deliverables positions this work as a critical enabler of high-quality outcomes and compliance. Future enhancements could include incorporating machine learning models for risk prediction or optimizing global submission workflows.