



# Improving Data and Analysis Operations through data science and technology innovation

Tiepu Liu, MD, PhD  
President, Global Biometrics



# Outline

- Data and Analysis Operations
  - Needs
  - Challenges
- How Data Science Services & Technology Help?
  - Data warehouse
  - Data standardization
  - Analysis and reporting

# Data and Analysis Operations

## Data Management

### Study Setup

- CRF design
- data management plan
- data validation
- database design and programming
- EDC site support

### Study Conduct

- Coding
- SAE reconciliation
- data transfer
- ongoing review

### Study Close Out

- DBL, archival

### EDC Systems

- Medidata Rave
- Oracle Inform
- Medrio
- OpenClinica
- Veeva

## Biostatistics

- Power / Sample Size Estimation/ Protocol Development
- Randomization
- Statistical Analysis Plan
- TLF Shells
- Interim Analysis / DMC Support
- Statistical Modeling and Simulation
- Clinical Study Report
- Statistical Report
- Annual Report (DSUR and PSUR)
- Observational Studies, RWE, and HEOR and Payer Analysis
- Publications

## Statistical Programming

- CRF, SAP Review and Inputs
- SDTM Data Package Preparation
- ADaM Data Package Preparation
- PK/PD Dataset Preparation
- Integrated SDTM/ADaM Data Packages (ISS/ISE/ISI) Preparation
- Office of Scientific Investigation (clinsite.xpt, define.xml, reviewers guide and OSI listings)
- Patient Profiles
- Data Exploration and Programming Automation

## Data Science and Technology

- Data server and security, cloud storage and computing infrastructure
- Middleware platform
- System connector and dynamic data integration
- RWD, Data standardization, integration and repository
- Data model, visualization, and dashboard reporting
- AI/ML, system and process automation and big data analytics
- Python, R tools and library

## FSP

### FSP Models

- Staff Augmentation Model
- Managed Staffing Model
- Full Outsourcing FSP Model
- Hybrid Model

### Experience

- Multiple major FSP providers, unitized model
- Selected Clients, FTE model
- Inhouse staff and subcontractors
- DM, Statistics, and Programming
- Safety and Regulatory, mostly offshore

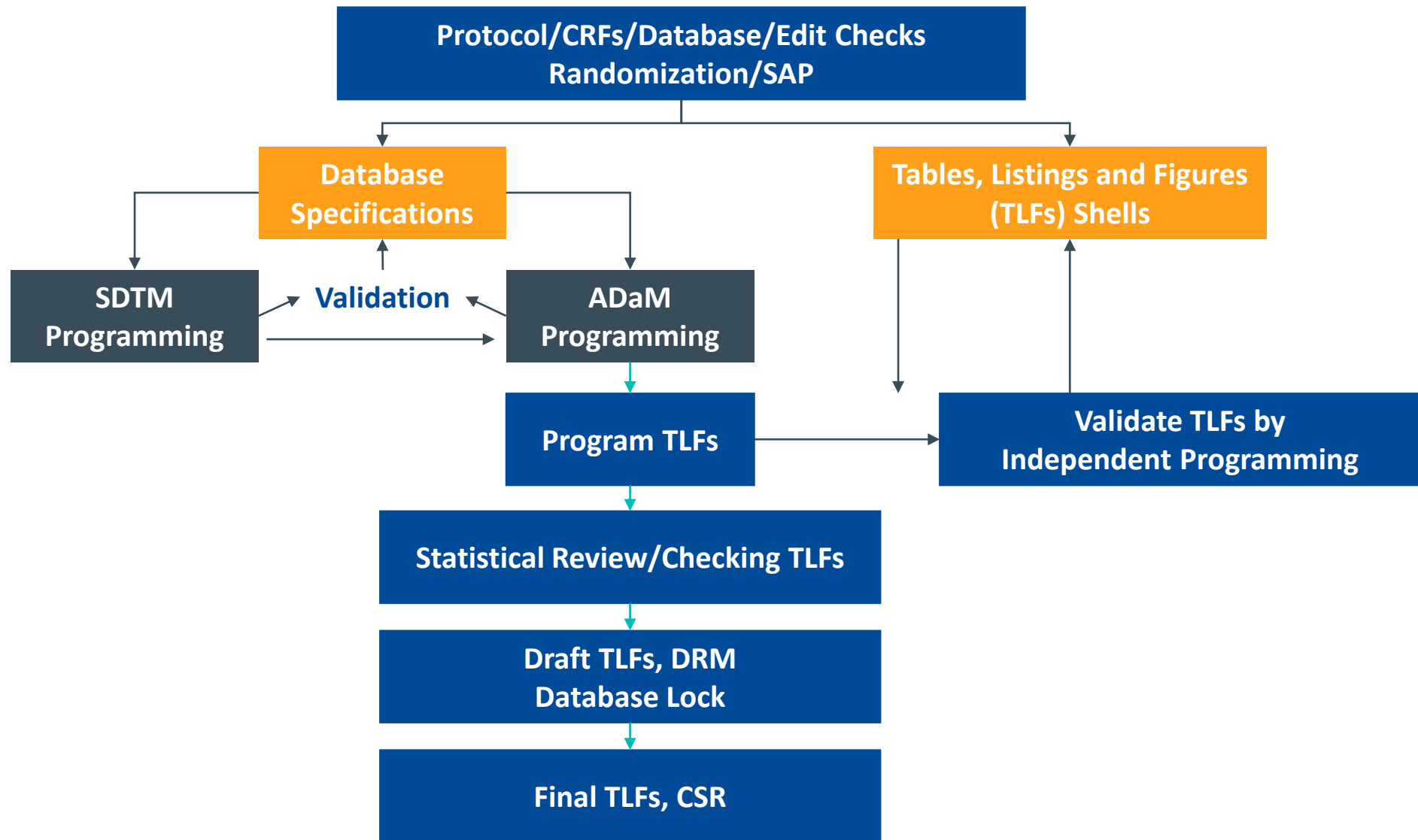
### FSP Management

- Internal dedicated FSP group
- Engagement team, account managers
- Resource managers
- HR recruiter team
- Database and agency network

# Data Operations Challenges

- Data Quality
  - Data Management
  - Risk Management
- Complexity
  - Tedious, manual
- Time, Cost and Efficiency

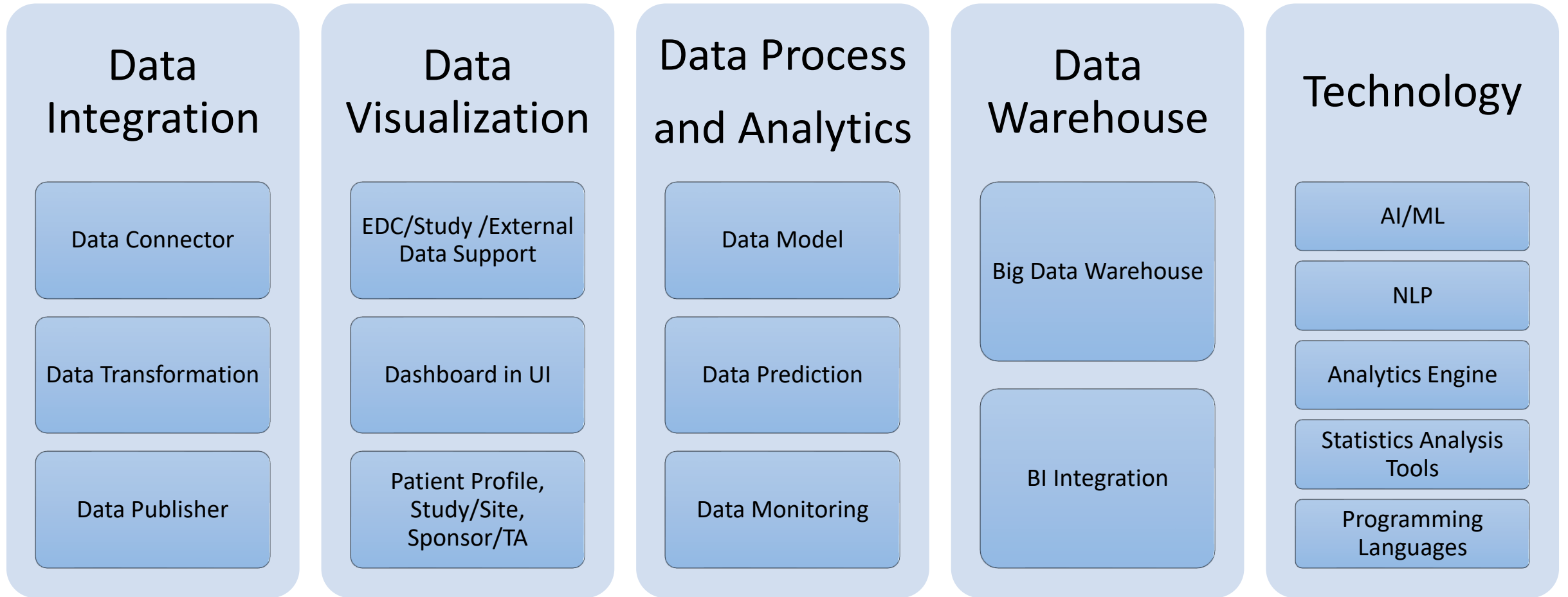
# Data and Analysis Process and Challenges



# How Data Science Services & Technology Help?

- Data Connection
  - Automatic connect (API) and dynamic update
- Data warehouse and standardization
  - CDISC and CDASH
  - SEND/SDTM/ADaM
- Analytics
  - Data visualization
  - Analysis and Reporting
    - TLFs

# General Data Science Services & Technology



# Software & Technology Portfolio

- Biometrics

- SAS for statistical programming
- SAS Libraries, Python/R packages
- Statistical planning and analysis systems/packages



- Data Science

- Data Capture and Process - Python, R, SAS, Java
- Data Visualization & Reporting – PowerBI, Spotfire, Tableau, Matplotlib, Plotly
- Data Warehouse/Data Lake – Azure Data Warehouse & Data Lake, Snowflake, Amazon Redshift, Apache HIVE
- Data Analytics - Apache Spark, Azure Synapse Analytics
- AI & ML - Scikit-Learn, Pandas, AutoML, NLP, etc.



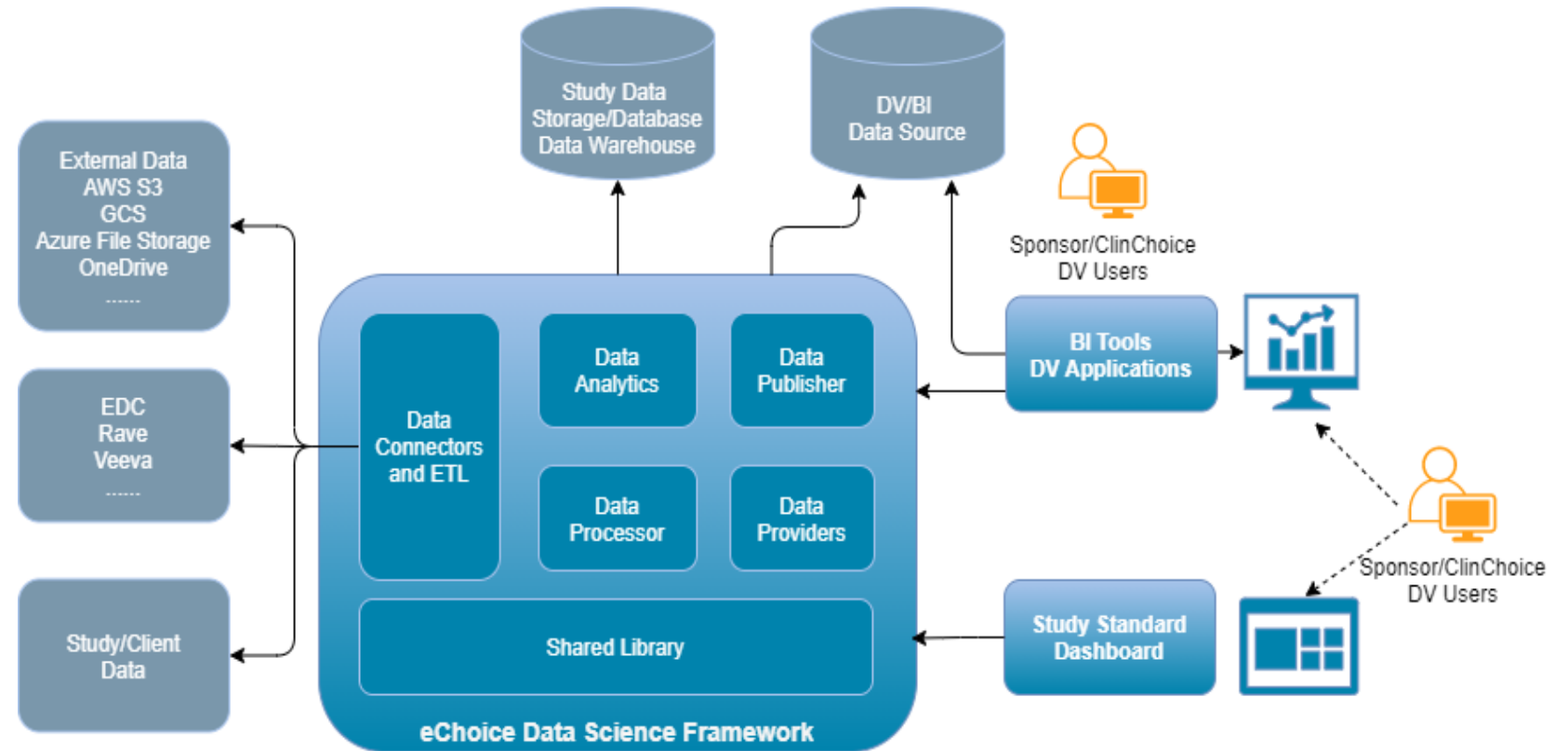
# **Data Science Services & Technology Applications**

## **Data Warehouse and Standardization**



# Data Process Framework

- Integrated Data Science Framework
- Consolidated Data Sources
- Process Pipeline and Automation
- Security & Compliance
- Integrated Tools



# Data Standardization

- eCRF Library and Builder
- SDTM Mapping
- aCRF Annotation/Bookmark Generator
  - Use DM Spec as Input
  - Automatically Locate/Insert Annotation
  - Support Customized Annotation styles
  - Support Complex Bookmark Structure

The screenshot displays the 'aCRF PDF' application interface. At the top, there are four tabs: 'Generate Annotations', 'Generate Bookmarks', 'Export Annotations/Bookmarks', and 'Import Annotations/Bookmarks'. Below the tabs, the 'aCRF PDF' field is set to 'C:\aCRF\Blank CRF.pdf' and the 'Spec File' is 'C:\aCRF\jarrett\spec.xlsx'. The 'Options' section is divided into two panels. The left panel, 'Import Annotations/Bookmarks', has 'Include Annotations' and 'Include Bookmarks' checked, and 'Keep Old Annotations & Bookmarks?' unchecked. The right panel, 'Generate Bookmarks', has 'Root Node' set to 'Study Name -123', 'Domains Node' set to 'By Domains', 'Visits Node' set to 'By Visits', and 'Visit Sheet Prefix' set to 'Visit Structure'. There are also checkboxes for 'Use Form Name in Visit Sheet?' and 'Keep Old Bookmarks?'. Below the options, there are two tabs: 'DM = Demographics' (selected) and 'XM = Demographics Rescreened Subjects'. The 'Version 0.2: Unique' section shows 'Form: Subject Enrollment' and 'Generated On: 15 Sep 2021 13:36:53'. The 'DM.STUDYID' and 'XM.STUDYID' fields are both set to '904-TRM-101'. The 'Subject Number' field is 'NOT SUBMITTED'. The 'Site ID' field has 'DM.SITEID' and 'XM.SITEID' tabs. The 'Cohort' field has 'SUPPDM.QVAL where QNAM = COHORT' and 'SUPPXM.QVAL where QNAM = COHORT' tabs. The 'Subject ID' field has 'DM.SUBJID' and 'XM.SUBJID' tabs. The 'Was this subject a Screen Failure?' field is 'NOT SUBMITTED'. There are 'Yes' and 'No' radio buttons at the bottom right.

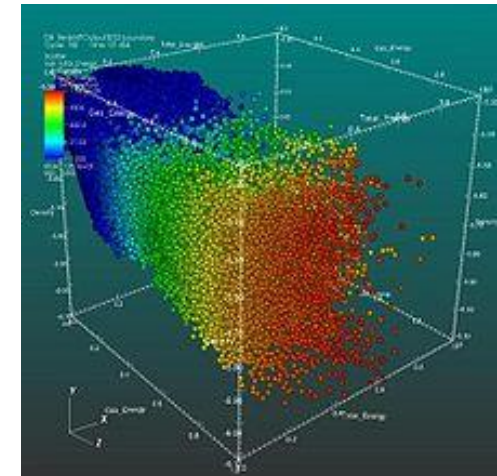
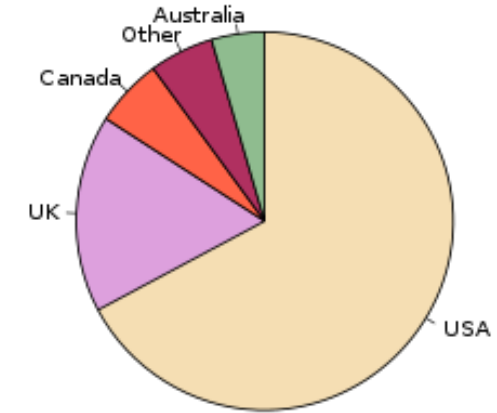
# **Data Science Services & Technology Applications**

## **Data Visualization**



# Data Visualization Applications

- Data visualization
  - Graphic representation of data,
  - Both an art and a science
- Better Analysis
  - Telling a story through data
  - Interactive & Customizable
  - Handling large amounts of data in a visual format
  - Finding errors/causes
  - Identifying patterns
  - Grasping the latest trends
- Signal Detection and Decision Making
  - Revealing previously unnoticed key points
  - Exploring business insights
  - Assisting decision making
  - Quick action



# Data Visualization Features

- General Features
  - Integrated Framework – EDC, Safety, Cloud Storage Excel Files, Raw/SDTM/ADaM Datasets
  - Full Automation, Near Real Time Data
  - Interactive & Customizable
  - Data Integrity - Finding errors/causes
  - Assisting Decision Making
  - Better Budget Control
- Enhanced Features
  - Complicated charts with Python/R
  - Support variable mappings to apply multiple studies – efficiency
  - Powerful navigation - Drill down & Drill up capability
  - User defined filters
  - User defined charts
  - Reporting capabilities
  - Alerts

# Data Visualization in Clinical Development

## Patient Profile

- Individual View
- Group View
- Customized View
- Reports

## Project Summary

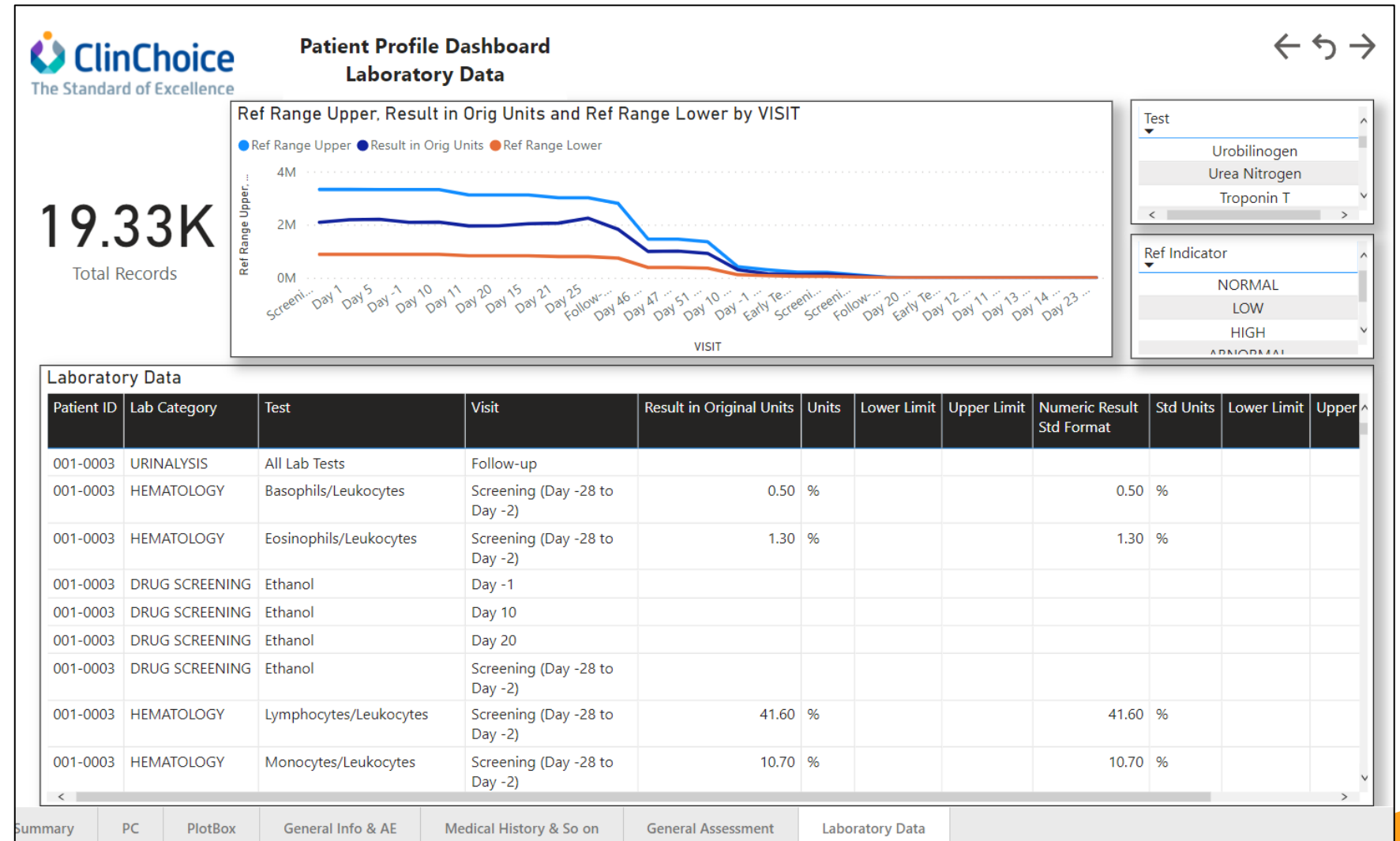
- PD Report
- DM Metrics
- Summary View
- Group View
- Customized View

## Risk Based Monitoring (RBM)

- KPI
- KRI
- Reports

# Individual Patient Profile

- Individual View
  - DM
  - Disposition
  - EX
  - MH
  - VS
  - LAB
  - AE
- Drill Down
  - Figure
  - Table/Listing
- Customized View



# Dynamic Patient Profile

## Patient Profile Filter

### Patients with Rash AE in first 21 Days

- ☒ No  
☒ Yes

### Uveal Melanoma Stage at Initial Diagnosis

- ☒  
☒ I  
☒ II  
☒ III  
☒ IV

### Baseline ECOG Performance Category

- ☒  
☒ <=3

### Baseline LDH Category

- ☒  
☒ Above ULN  
☒ Below ULN

### Prior Anti-Cancer Radiotherapy Surgery

- ☒  
☒ Y

### Prior Anti-Cancer Systemic Therapy

- ☒  
☒ Y

### Prior Anti-Cancer Surgery

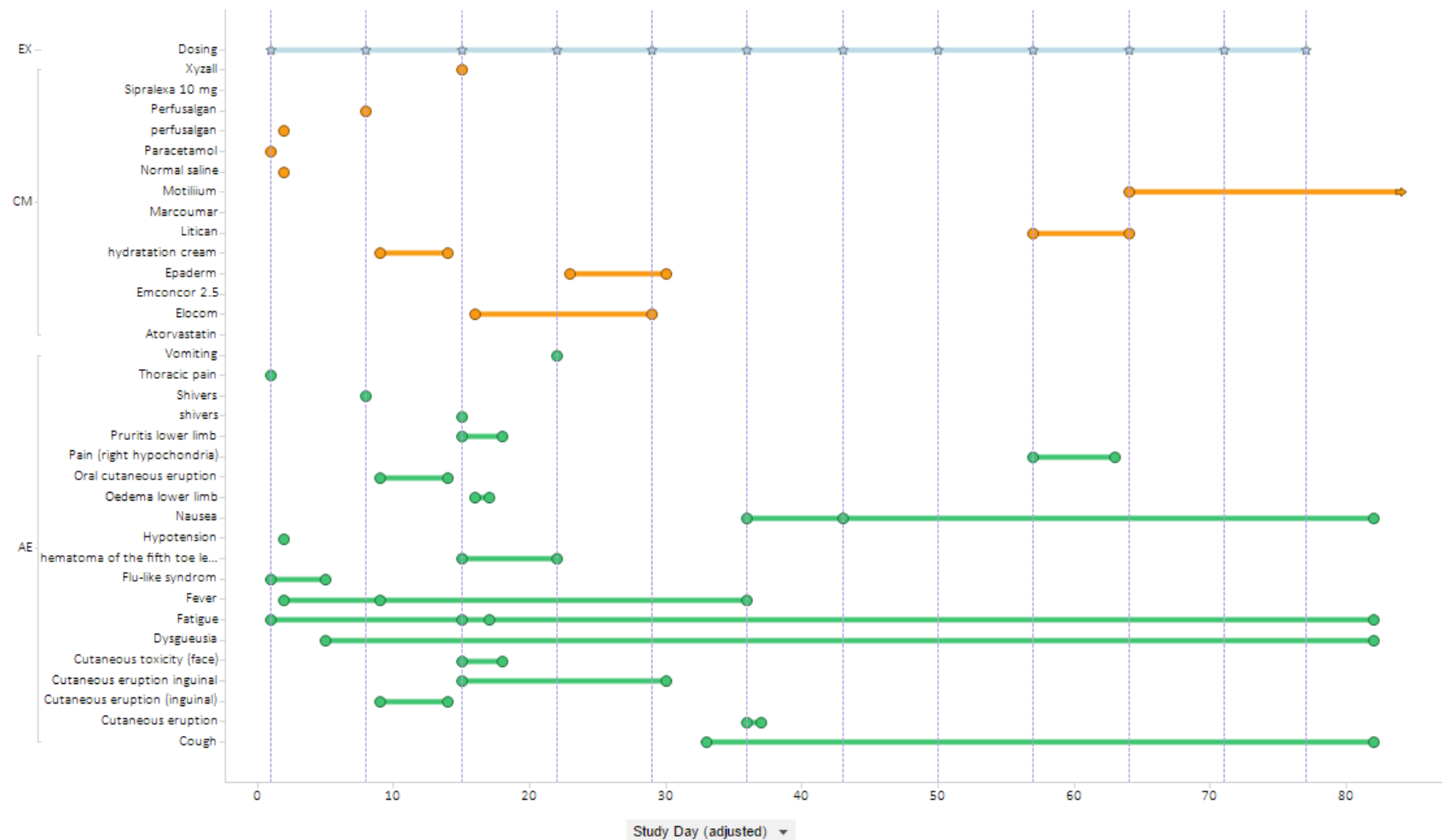
- ☒  
☒ Y

### Anti-cancer Treatment since Discontinuation

- ☒

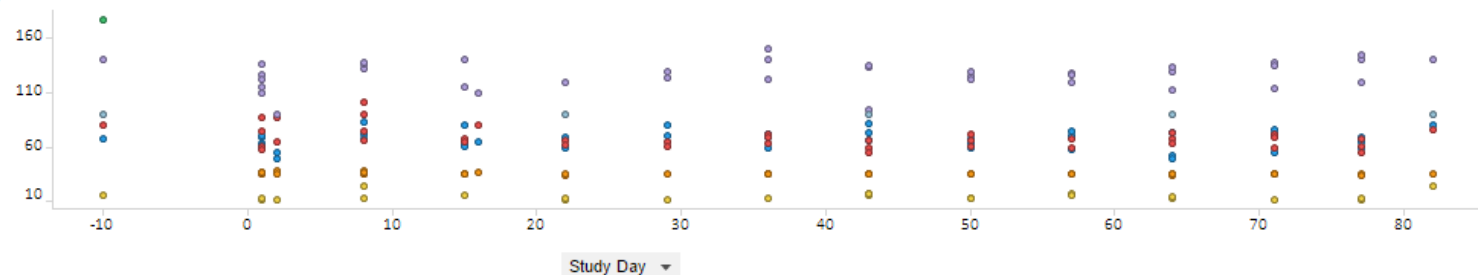
## Patient Profile

Domain Abbreviation PTCODE



## Vital Signs Result vs. Study Day

Numeric Result/...

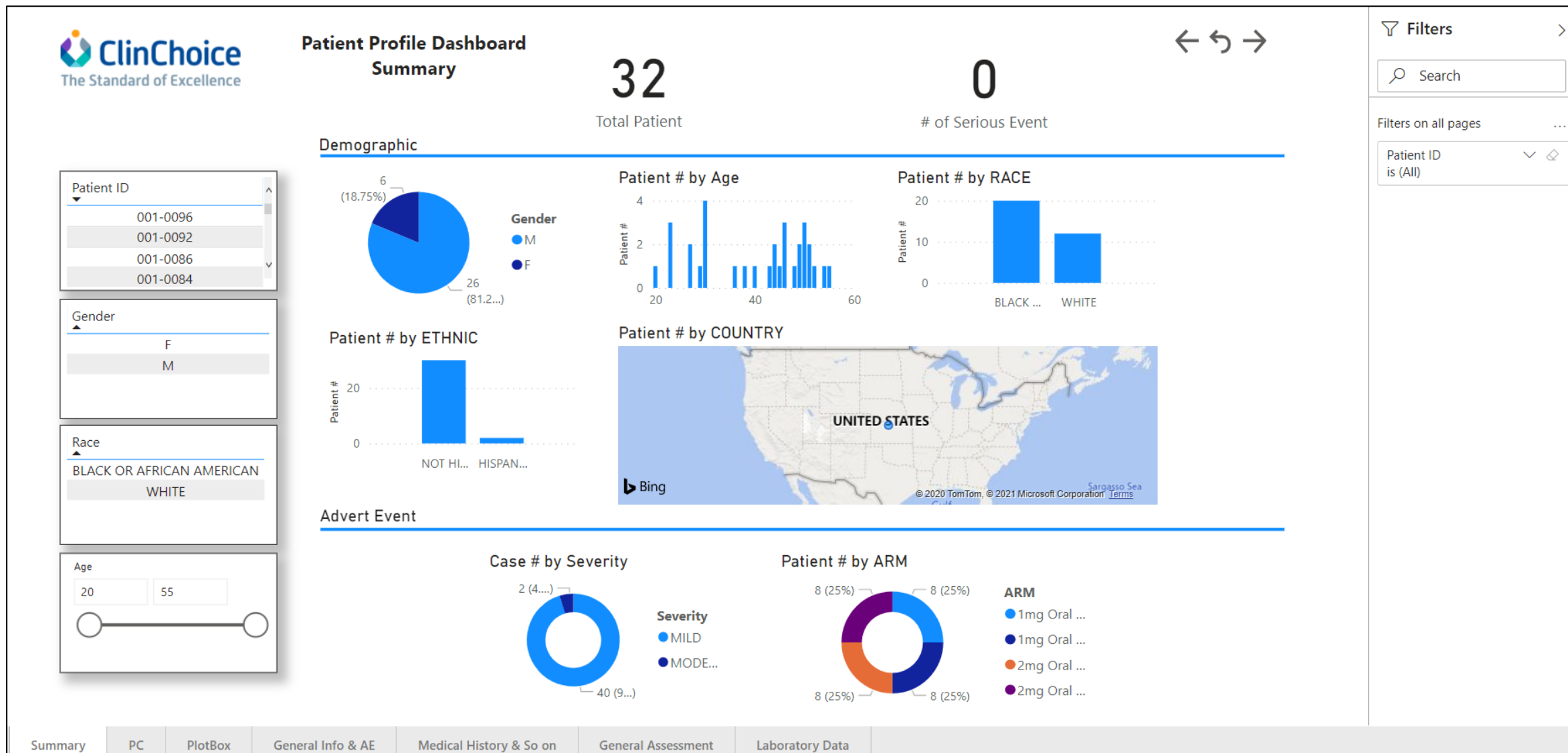


# **Data Science Services & Technology Applications**

## **Project Profile – Study Dashboard and RBM**



# Project Profile – Dynamic Status Dashboard



# Risk Based Monitoring Plan with KRI/KPI

- Risk Based Monitoring Implementation
  - Assessment - Collecting KRIs (Key Risk Indicator) / KPIs (Key Performance Indicator)
    - KRI/KPI Identification with Category/Sub-Category
      - Data Quality/Compliance/Safety/CRA Workload/Issue Management ...
    - KRI/KPI Level & Threshold
    - Level of Scrutiny
    - KRI/KPI Measure
    - Risk Control Options
  - Design - Determine Data Sources, Data Presentation, Data Process Flow and Communication
  - Implementation - Develop Monitoring Function/Flow on Data Science Platform and BI Tools
    - Integration to Various Data Sources
    - Integration to BI Tool
    - Data Process Flow with Automation
    - System Configuration
  - Continuous Improvement of Monitoring

# Risk Based Monitoring Approach

- Data Visualization for Data Quality and Risks with Frontend BI Tool
- Risk/Performance Assessment Implemented in Backend Middleware
- Communication with Alert/Reminder via Email/Message etc.
- Reporting and Auditing for Review
- Decision Support for Risk Control

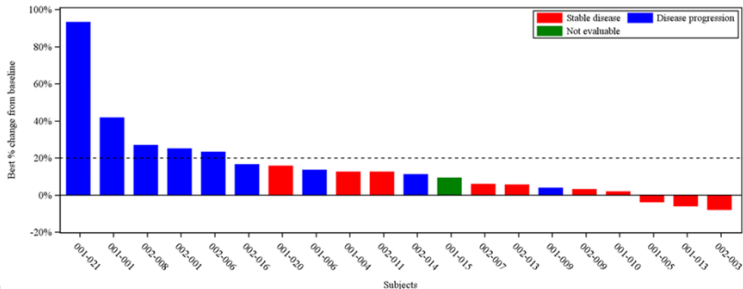
# **Data Science Services & Technology Applications**

## **Analysis and Reporting Efficiency Improvement**

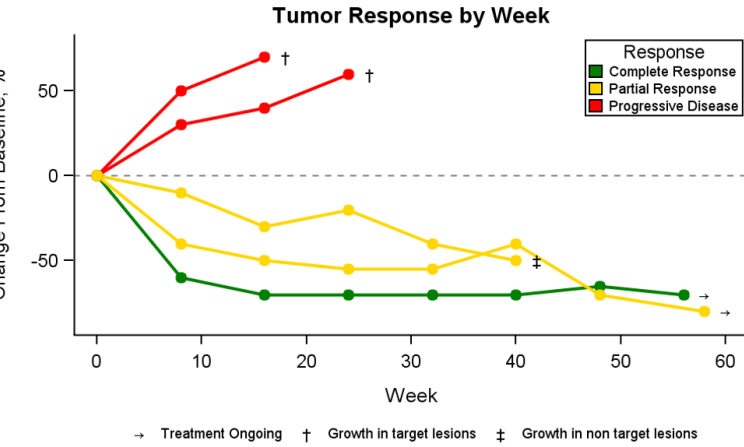
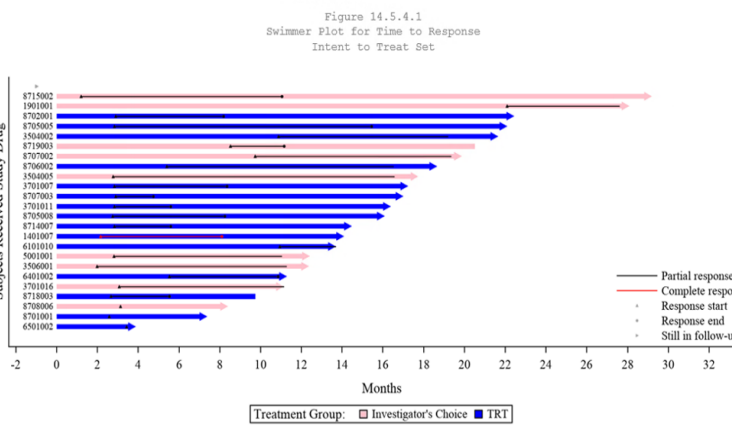
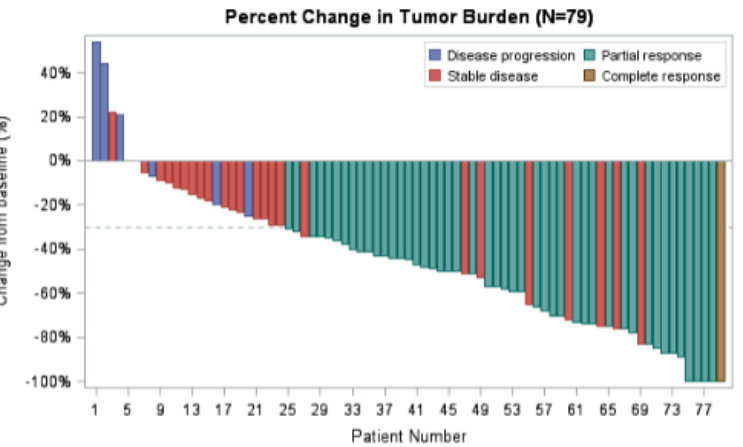
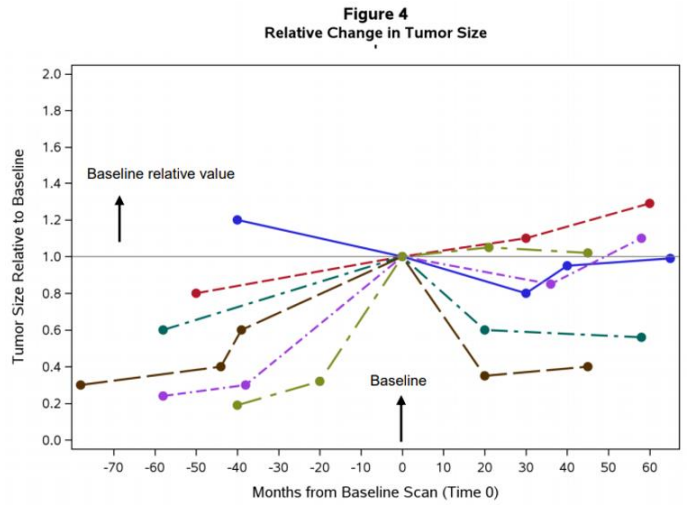
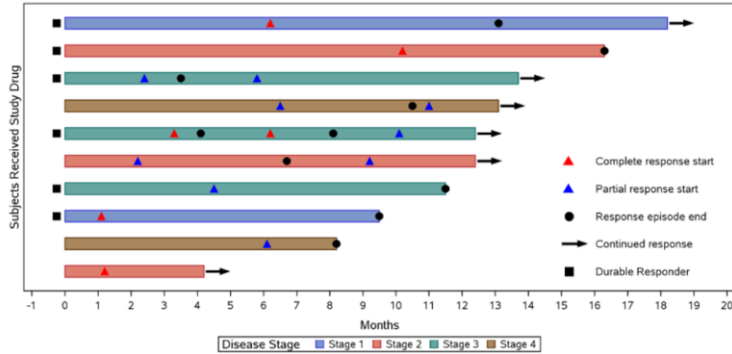


# Analysis Reporting Automation

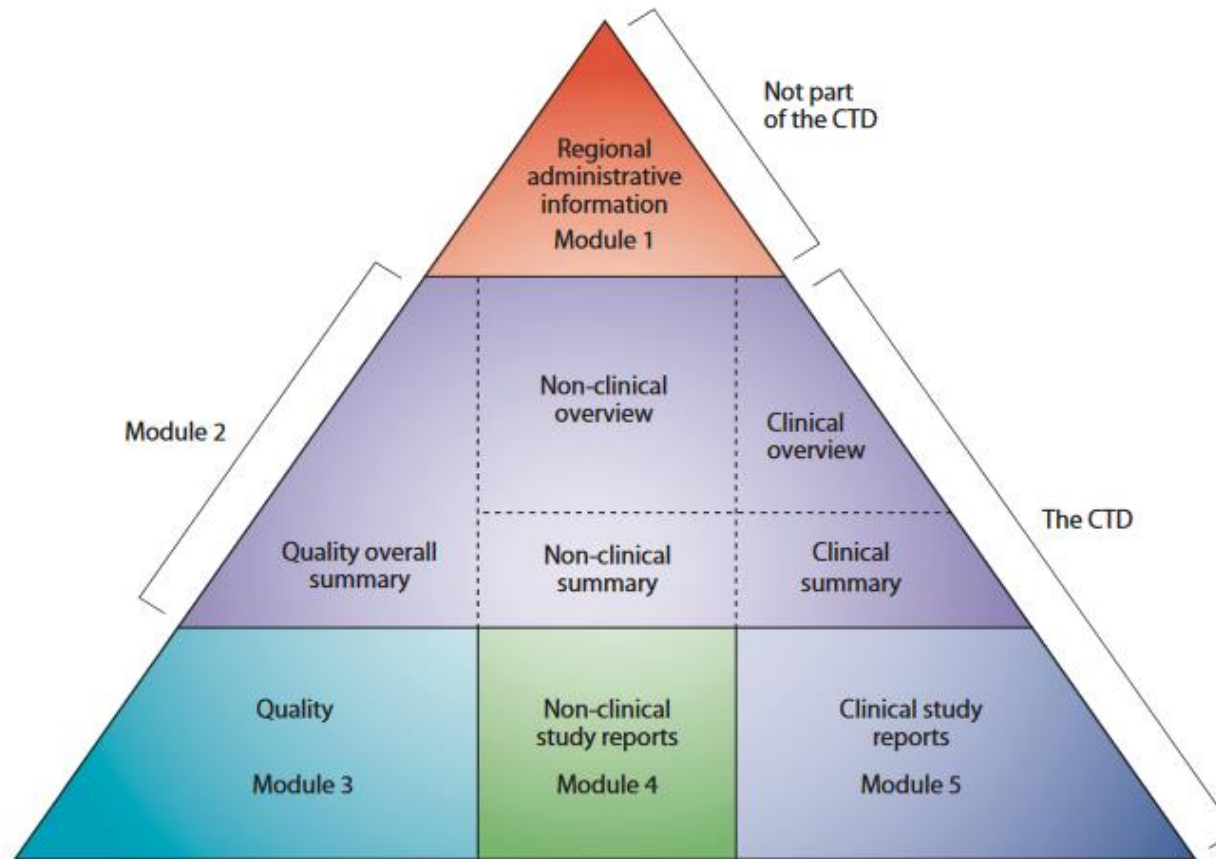
## Statistical Graphics – Waterfall, Swimmer, Spider Plots



Tumor size is measured as the sum of longest diameters or short axis of the target lesions according to RECIST version 1.1. Best percent change in target lesion size is the maximum percent reduction from baseline or the minimum percent increase from baseline.



# eData/eCTD Submission



The CTD triangle. The Common Technical Document is organized into five modules. Module 1 is region specific and modules 2, 3, 4 and 5 are intended to be common for all regions.

Folder Name	Folder Level	Description/Contents
📁 [module]	1	Refers to the eCTD module in which study data is being submitted. Name this folder m4 for nonclinical data and m5 for clinical data. Do not place files at this level.
📁 datasets	2	Resides within the module folder as the top-level folder for study data (nonclinical or clinical) being submitted for the specified module (m4 or m5). Do not place files at this level.
📁 [study]	3	Name this folder with the study identifier or analysis type performed (e.g., study123, iss, ise). Do not place files at this level.
📁 analysis	4	Contains folders for analysis datasets and software programs; arrange in designated level 6 subfolders. Do not place files at this level.
📁 adam	5	Contains subfolders for ADaM datasets and corresponding software programs. Do not place files at this level.
📁 datasets	6	Place ADaM datasets in this subfolder.
📁 split	7	Place any split ADaM datasets in this subfolder.
📁 programs	6	Place software programs for ADaM datasets, tables and figures in this subfolder.
📁 legacy	5	Contains legacy formatted analysis datasets and corresponding software programs. Do not place files at this level.
📁 datasets	6	Place legacy analysis datasets in this subfolder.
📁 split	7	Place split legacy analysis datasets in this subfolder.
📁 programs	6	Place software programs for legacy analysis datasets, tables and figures in this subfolder.
📁 misc	4	Place miscellaneous datasets that don't qualify as analysis, profile, or tabulation datasets in this subfolder. This subfolder was formerly named "listings".
📁 profiles	4	Place patient profiles in this subfolder.
📁 tabulations	4	Contains subfolders for tabulation datasets. Do not place files at this level.
📁 legacy	5	Place legacy (non-standardized) tabulation datasets in this folder.
📁 split	6	Place any split legacy tabulations datasets in this subfolder.
📁 sdtm	5	Place SDTM tabulation datasets in this subfolder. Should only be used in m5 for clinical data.
📁 split	6	Place any split SDTM files in this subfolder.
📁 send	5	Place SEND tabulation datasets in this subfolder. Should only be used in m4 for animal data.



# THANK YOU

**ClinChoice Inc.**

1300 Virginia Drive, Suite 408  
Fort Washington, PA 19034 USA  
Phone: 215-283-6035  
Fax: 215-283-6382  
Email: [info@clinchoice.com](mailto:info@clinchoice.com)

**ClinChoice Private Ltd.**

#729 8th Main  
11th Cross, J.P. Towers  
JP Nagar 3rd Phase  
Bangalore - 560 076  
Karnataka, India  
Phone: +91 80 2658 5551  
Email: [info@clinchoice.com](mailto:info@clinchoice.com)

**ClinChoice**

No. 420 Fenglin Road  
Floor 11  
Xuhui District  
Shanghai 430000, China  
Email: [bd@clinchoice.com](mailto:bd@clinchoice.com)

**ClinChoice Inc.**

26th Floor Wynsum Corporate Plaza  
22 F. Ortigas Jr. Road  
Ortigas Center  
Pasig City, - 1605  
Philippines  
Phone: +632 633-5176  
Email: [info@clinchoice.com](mailto:info@clinchoice.com)

**ClinChoice Limited**

Suite G48  
268 Bath Road, Slough, SL1 4DX  
Berkshire, United Kingdom  
Phone: +44-(0)-1628-566121  
Email: [europe@clinchoice.com](mailto:europe@clinchoice.com)

**ClinChoice K.K.**

THE GATE NIHONBASHI Bld. 3FHigashi  
Nihonbashi 3-4-1, Chuo-ku  
Tokyo 104-0031 Japan  
Phone: 81-3-6228-7360  
Email: [japan@clinchoice.com](mailto:japan@clinchoice.com)

**ClinChoice LLC**

11 Kalents Street, 0033  
Yerevan, Armenia  
Phone: 374-60-541515  
Email: [europe@clinchoice.com](mailto:europe@clinchoice.com)

**ClinChoice Canada Inc.**

2 Robert Speck Parkway, Suite 750  
Mississauga, Ontario L4Z 1H8, Canada  
Phone: 215.283.6035  
Email: [info@clinchoice.com](mailto:info@clinchoice.com)