WITH STANDARDS – UNLOCK THE POWER OF DATA

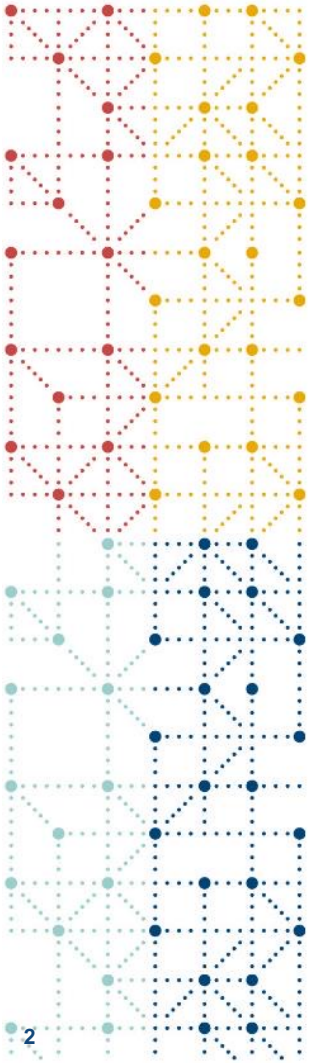
cdisc

2022
US
INTERCHANGE
26-27 OCTOBER | AUSTIN

Medical Imaging Data Standards, Automation & Analysis

OCTOBER 2022

Shyam Banuprakash and Kim Nguyen, Data Science, Clario



Agenda

1. Clario – Introduction
2. Data Standards
3. Clinical Data Automation and Management System
4. Machine Learning Case Study

Who We Are

CLARIO.

OUR PURPOSE

To transform lives by unlocking better evidence

AT A GLANCE

50
almost 50 years' experience

70%
of all FDA approvals (2019–2020)

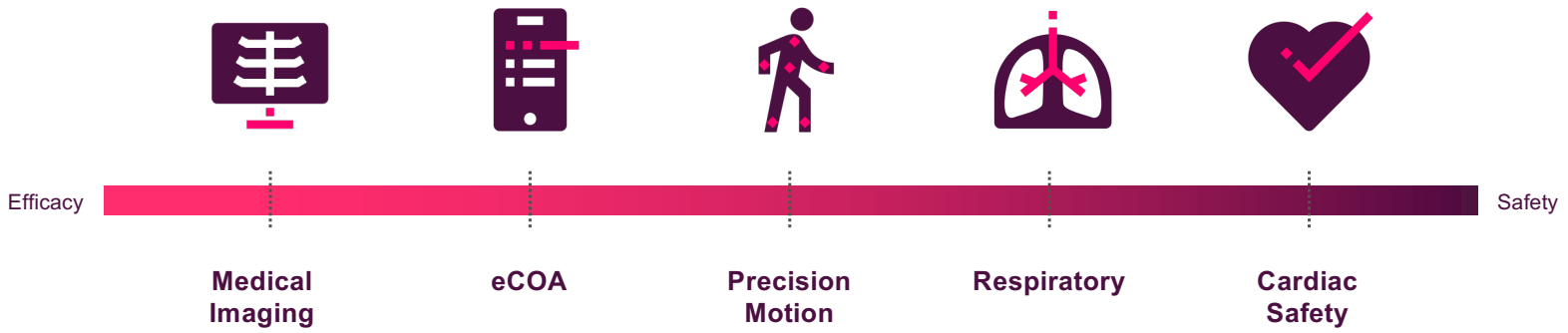
30
facilities in 9 countries across Europe, North America and Asia Pacific

19k
clinical trials

870
regulatory approvals

24/7
customer and patient support

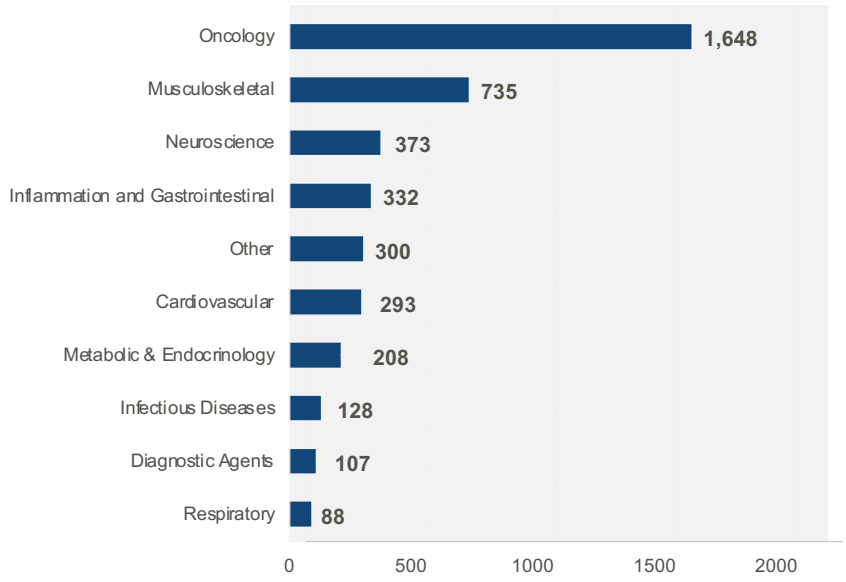
The Broadest Endpoint Technology Platform



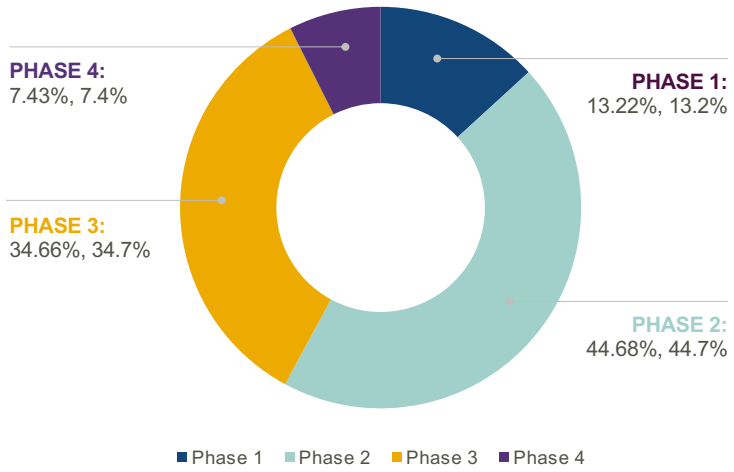
Medical Imaging Experience

4,200+ studies

Number of Studies by Therapeutic Area



Number of Studies by Phase



Oncology Expertise

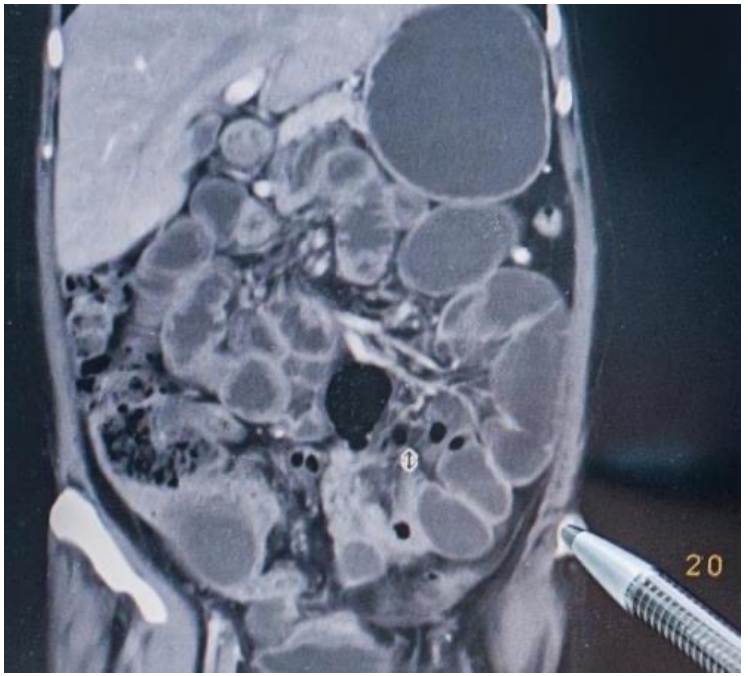
Clario is the world leader in oncology imaging, helping to **assess efficacy** and **safety** in thousands of trials

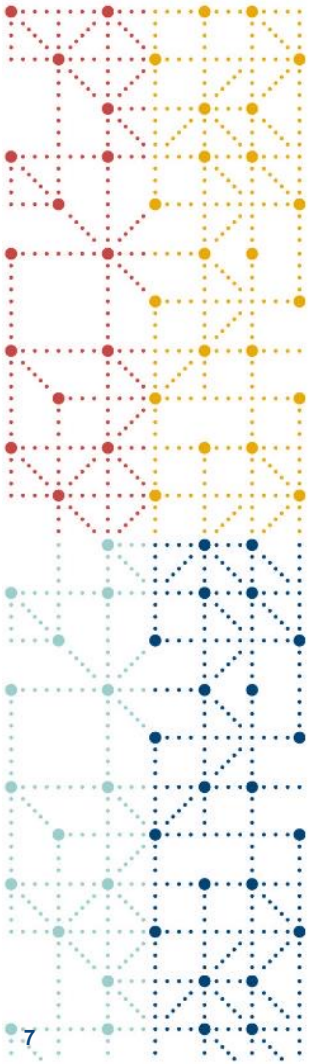
1,700+ oncology studies

220+ FDA-approved oncology drugs

Readers > dedicated and in-house

Leadership in breast, lung, and lymphoma Medical Imaging expertise supporting multiple tumor types in Phase I-IV studies





Data Driven Culture

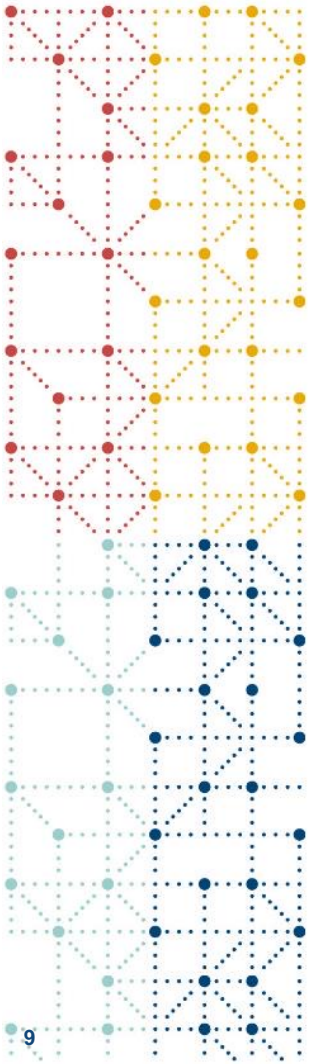
Introduction

Data is the lifeblood of organizations.

Insights from data are the key to success.

Basic data analytics can be straightforward, but companies can be overwhelmed as they progress.

Building a data-driven culture and trust around data analysis is essential for long-term success.



Data Standards

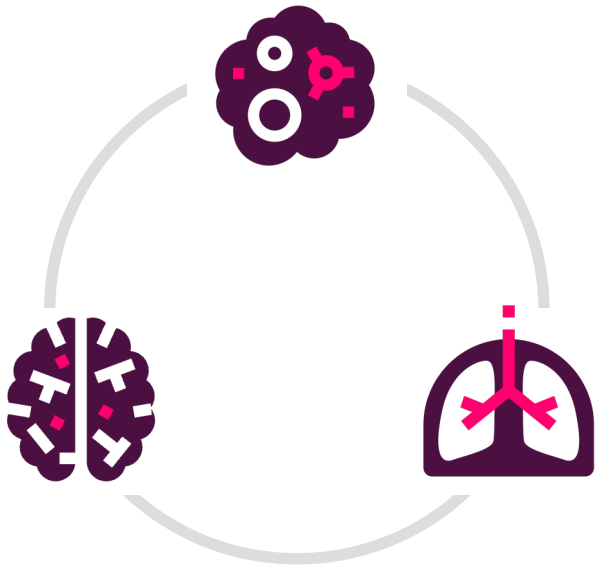
Medical Imaging

Medical Imaging

- **Medical Imaging forms a significant part of most clinical trials**
- **Used to generate study endpoints**
  - Primary, secondary or exploratory study endpoints
- **Help determine the safety and efficacy of the treatment.**
- **Typically, imaging is assessed by using established response criteria such as RECIST 1.1.**

Medical Imaging Data Standards

- **Study Data Tabulation Model (SDTM) implementation guide**
  - Study data regulatory submission guidelines
  - Provides information on the different domains, variables as well as formats to be used when submitting data
- **Oncology domains in SDTM**
- **TU**
  - Uniquely identifies tumors, lesions or locations of interest
- **TR**
  - Represents quantitative measurements and/or qualitative assessments of tumors, lesions or locations of interest
- **RS**
  - Assessment of disease response to therapy, or clinical classification based on published criteria



TU Domain in Oncology Studies

tu.xpt

Row	STUDYID	DOMAIN	USUBJID	TUSEQ	TULNKID	TUTESTCD	TUTEST	TUORRES	TUSTRESC	TULOC	TULAT	TUMETHOD	TUNAM	TUEVAL	TUEVALID	VISITNUM	VISIT	TUDTC	TUDY
1	ABC	TU	55555	1	R1-T01	TUMIDENT	Tumor Identification	TARGET	TARGET	CERVICAL LYMPH NODE	LEFT	MRI	ACE IMAGING	INDEPENDENT ASSESSOR	RADIOLOGIST 1	10	SCREEN	2010-01-02	-2
2	ABC	TU	55555	2	R1-T02	TUMIDENT	Tumor Identification	TARGET	TARGET	LIVER		CT SCAN	ACE IMAGING	INDEPENDENT ASSESSOR	RADIOLOGIST 1	10	SCREEN	2010-01-01	-3
3	ABC	TU	55555	3	R1-T03	TUMIDENT	Tumor Identification	TARGET	TARGET	THYROID GLAND	RIGHT	CT SCAN	ACE IMAGING	INDEPENDENT ASSESSOR	RADIOLOGIST 1	10	SCREEN	2010-01-01	-3
4	ABC	TU	55555	4	R1-NT01	TUMIDENT	Tumor Identification	NON-TARGET	NON-TARGET	KIDNEY	RIGHT	CT SCAN	ACE IMAGING	INDEPENDENT ASSESSOR	RADIOLOGIST 1	10	SCREEN	2010-01-01	-3
5	ABC	TU	55555	5	R1-NT02	TUMIDENT	Tumor Identification	NON-TARGET	NON-TARGET	CEREBELLUM	RIGHT	MRI	ACE IMAGING	INDEPENDENT ASSESSOR	RADIOLOGIST 1	10	SCREEN	2010-01-02	-2
6	ABC	TU	55555	6	R1-NEW01	TUMIDENT	Tumor Identification	NEW	NEW	LUNG		CT SCAN	ACE IMAGING	INDEPENDENT ASSESSOR	RADIOLOGIST 1	40	WEEK 6	2010-02-20	48
7	ABC	TU	55555	7	R1-NEW02	TUMIDENT	Tumor Identification	NEW	NEW	CEREBELLUM	LEFT	MRI	ACE IMAGING	INDEPENDENT ASSESSOR	RADIOLOGIST 1	60	WEEK 12	2010-04-02	88

In a RECIST study, tumors/lesions of interest are categorized as target, non-target and new tumors

The TU domain will capture

- Subject identifier
- Role of the evaluator
- LinkID used to identify the tumor
- Location of the tumor
- Method used to identify the tumor
- Type of tumor (target, non-target, new)

SDTM is the Recommended Data Standard Used for Submitting Clinical Trial Data to the FDA

TR Domain

tr.xpt

Row	STUDYID	DOMAIN	USUBJID	TRSEQ	TRGRPID	TRLNKGRP	TRLNKID	TRTESTCD	TRTEST	TRORRES	TRORRESU	TRSTRESC	TRSTRESN	TRSTRESU	TRNAM	TRMETHOD	TREVAL	TREVALID	VISITNUM	VISIT	TRDTC	TRDY
1	ABC	TR	55555	1	TARGET	A1	R1-T01	DIAMETER	Diameter	20	mm	20	20	mm	ACE IMAGING	MRI	INDEPENDENT ASSESSOR	RADIOLOGIST 1	10	SCREEN	2010-01-02	-2
2	ABC	TR	55555	2	TARGET	A1	R1-T02	DIAMETER	Diameter	15	mm	15	15	mm	ACE IMAGING	CT SCAN	INDEPENDENT ASSESSOR	RADIOLOGIST 1	10	SCREEN	2010-01-01	-3
3	ABC	TR	55555	3	TARGET	A1	R1-T03	DIAMETER	Diameter	15	mm	15	15	mm	ACE IMAGING	CT SCAN	INDEPENDENT ASSESSOR	RADIOLOGIST 1	10	SCREEN	2010-01-01	-3
4	ABC	TR	55555	4	TARGET	A1		SUMDIAM	Sum of Diameter	50	mm	50	50	mm	ACE IMAGING		INDEPENDENT ASSESSOR	RADIOLOGIST 1	10	SCREEN		
5	ABC	TR	55555	5	TARGET	A1		SUMNLNLD	Sum Diameters of Non Lymph Node Tumors	30	mm	30	30	mm	ACE IMAGING		INDEPENDENT ASSESSOR	RADIOLOGIST 1	10	SCREEN		
6	ABC	TR	55555	6	NON-TARGET	A1	R1-NT01	TUMSTATE	Tumor State	PRESENT		PRESENT			ACE IMAGING	CT SCAN	INDEPENDENT ASSESSOR	RADIOLOGIST 1	10	SCREEN	2010-01-02	-2

The results (quantitative/qualitative) of the tumors/lesions identified in the TU domain are reported within the TR domain. The TR domain will contain

- Subject identifier
- Role of the evaluator
- LinkID used to link the records to the tumors reported in the TU dataset
- Method used to identify the tumor
- Test used to obtain the measurement or finding
- Results of the test

RS Domain

rs.xpt

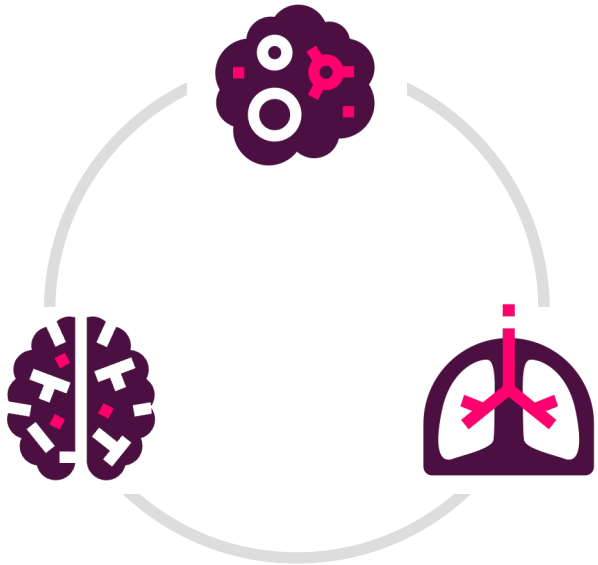
Row	STUDYID	DOMAIN	USUBJID	RSSEQ	RSLNKGRP	RSTESTCD	RSTEST	RSCAT	RSORRES	RSSTRESC	RSEVAL	VISITNUM	VISIT	RSDTC	RSDY
1	ABC	RS	44444	1		TRGRESP	Target Response	RECIST 1.1	PR	PR	INVESTIGATOR	40	WEEK 6	2010-02-18	46
2	ABC	RS	44444	2		NTRGRESP	Non-target Response	RECIST 1.1	SD	SD	INVESTIGATOR	40	WEEK 6	2010-02-18	46
3	ABC	RS	44444	3	A2	OVRLRESP	Overall Response	RECIST 1.1	PR	PR	INVESTIGATOR	40	WEEK 6	2010-02-18	46
4	ABC	RS	44444	4		TRGRESP	Target Response	RECIST 1.1	NE	NE	INVESTIGATOR	60	WEEK 12	2010-04-02	88
5	ABC	RS	44444	5		NTRGRESP	Non-target Response	RECIST 1.1	NE	NE	INVESTIGATOR	60	WEEK 12	2010-04-02	88
6	ABC	RS	44444	6		SYMPTDTR	Symptomatic Deterioration	PROTOCOL DEFINED RESPONSE CRITERIA	Pleural Effusion	PD	INVESTIGATOR	60	WEEK 12	2010-04-02	88
7	ABC	RS	44444	7	A3	OVRLRESP	Overall Response	PROTOCOL DEFINED RESPONSE CRITERIA	PD	PD	INVESTIGATOR	60	WEEK 12	2010-04-02	88

The results of the response assessment that might have been collected or calculated based on tumors/lesions identified in the TU domain and their results in the TR domain, are reported within the RS domain. The RS domain will contain

- Subject identifier
- Role of the evaluator
- Name of the response assessment (target/non-target/new response, date of progression, date of first response etc)
- Results of the response assessment

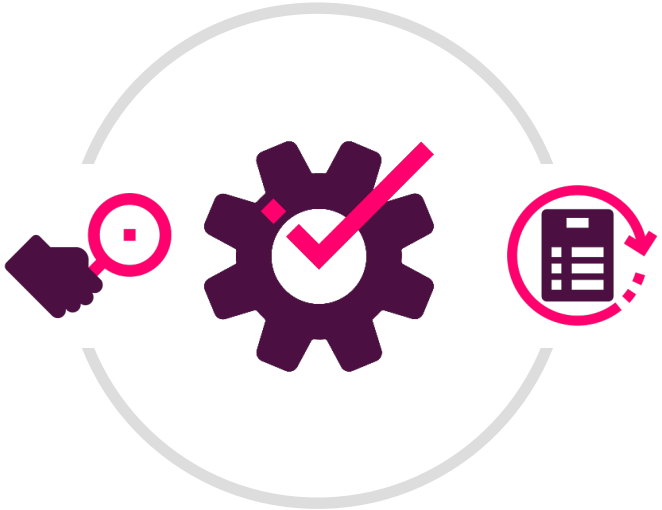
Challenges

- **Build databases to fit SDTM format and accommodate trial/criteria specificities**
- **Collection of data points in tumor evaluations for complex criteria**
  - Criteria vary in complexity and protocol requirements may be different
- **Multiple flavors of data standards to map to**
- **Perform data standardization and automation at scale**
- **Clario is managing the imaging endpoints for more than 500 active oncology studies**



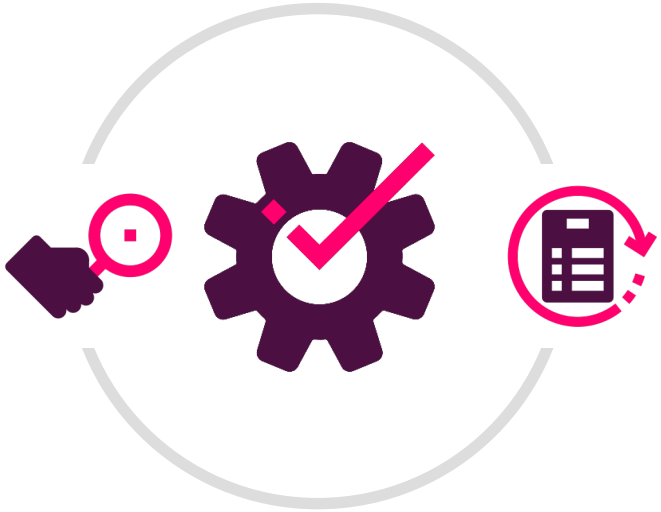
Solution Overview

- **Define a Clario data standard for Imaging RECIST 1.1 studies**
  - Based on the SDTM Guidelines
- **Build a system and workflow to standardize all studies**
- **Automate the entire workflow**
  - Automate data cleaning
  - Automate data analysis
  - Enable predictive analysis

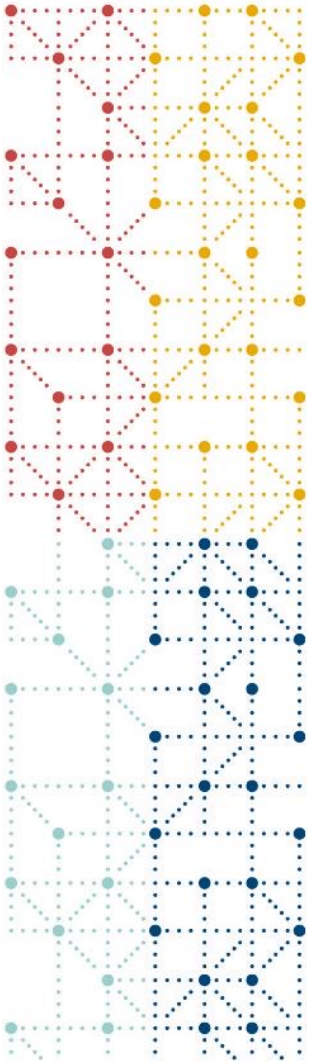


Clario's Standard

- Clario data standard is based on SDTM v3.4 but also accommodates different flavors along with the standard
- Clario's review of the data across different studies and clients have shown the same data point can be represented in different ways. For example, LNKID (or the lesion label) has been reported in multiple ways across the industry (Target lesions can be reported as T1, T01, R1-T01, R1T1 etc).
- Clario's approach will accommodate all the types along with its standard.
- Clario specifications and programming will not only generate a standard output (v3.4 SDTMIMG) but can also produce the specific types requested at the same time.



Variable Name	*Variable Label*	*Type*	*SAS Format*	*Controlled Terms, Codelist or Format***	*Description*	*Core*	*Default Value*	Drop (Y)	Custom Derivation_Method	CLARIO_STANDARD_OUTPUT
TULNKID	Link ID	Char	$8		Identifier used to link the assessment result records to the individual tumor identification record in TU domain.	Exp			TARGET: type_x, NON-TARGET: type_x, NEW: type_x, SPLIT: type_x or NA, MERGE: type_x or NA, No Disease: type_x or NA	TARGET: type_6, NON-TARGET: type_6, NEW: type_11, SPLIT: type_5, MERGE: type_6

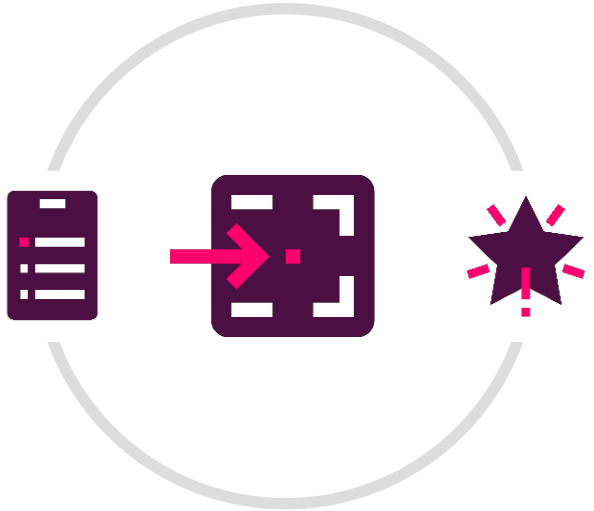


CDAMS

Clinical Data Automation & Management System

CDAMS – Clinical Data Automation & Management System

- Enable Clario to map all studies into a common set of standards
- Automate the workflow and the data cleaning process
- Provide users a single unified interface
  - To review and resolve the data discrepancies
- Enable real-time data sharing
- Enable real-time data analysis
- Enable prediction applications

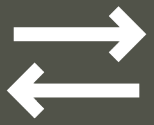


CDAMS Objectives

- Source System Independent
- Data Quality Improvement
- Time/Cost Savings
- Backward Compatible
- Enable the Creation of a "Clario Data Standard"
- Regulatory/ Compliance — Audit Ready System
- Enhanced Patient Reporting
- Real-time Data Sharing Capabilities

CDAMS Modules

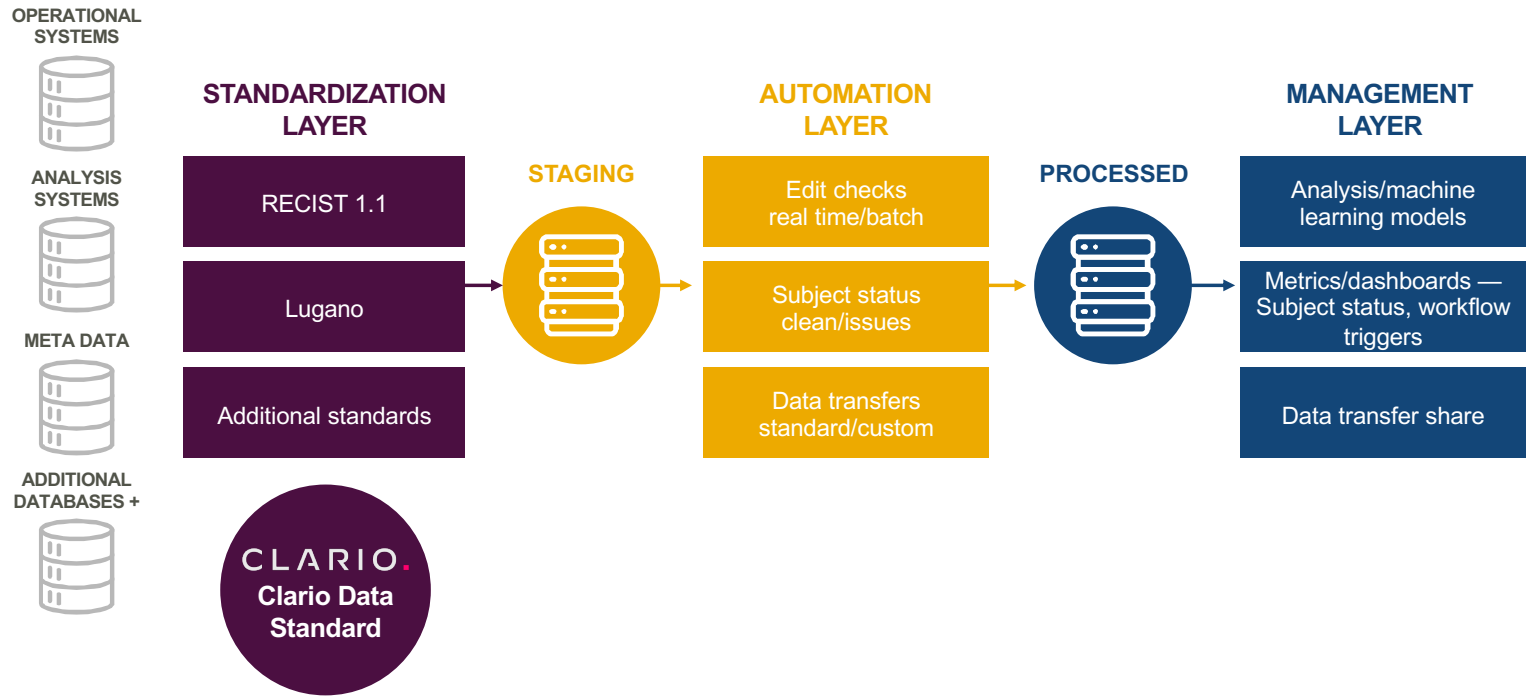
Data Cleaning Module



Data Transfer Standardization & Automation Module

Data Analysis, Machine Learning & Reporting Module

Application Architecture



Sponsor & Subject Reporting

Data Transfer

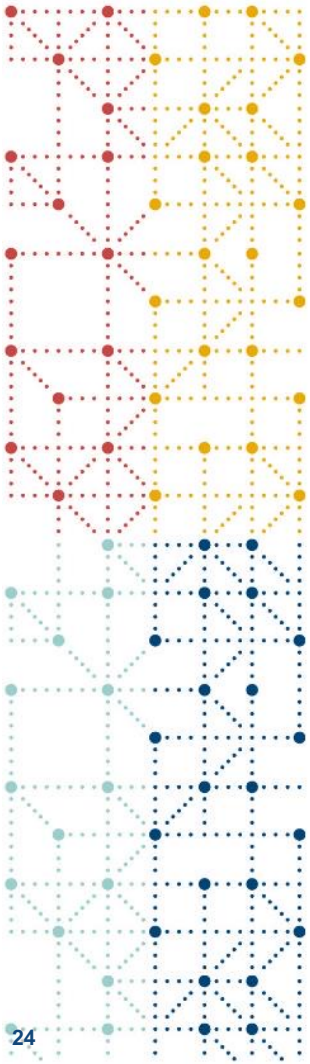
- Automate transfer
- Data share capabilities

Reporting

- Patient data reporting
- Near real-time view with controlled user access

Clinical Impact

- Adjudication monitoring
- Reader selection model
- Reader variability assessment automation



Data Analysis

Data Analysis

Standard Reporting

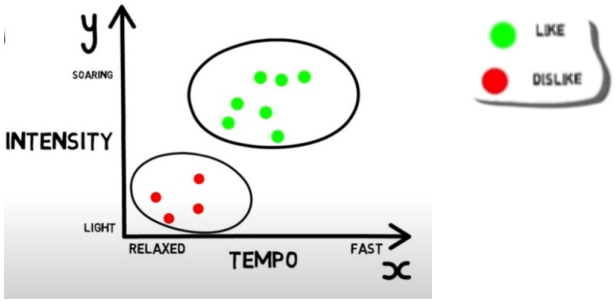
Predictive Analysis

Identifying Predictors to Influence Outcomes

What is Machine Learning?

Data Variables

1. Tempo
2. Genre
3. Gender of Vocalist
4. Sound intensity
5. Length of Song



How would you define rules to also incorporate the Genre, Vocalist, and/or Length of the Song?

How do we know which variable is relevant to the prediction?

How do we scale and test our predictions?

Benefits of Predictive Modeling

- Better understand the features that predict specific outcomes
- Support early detection of outcomes
- Identify subjects who are more likely to get specific outcomes
- Enhance and guide patient selection
- Better inform clinical study design

Case Study – Alzheimer's Cognitive Disease Prediction

Background

Alzheimer's disease (AD) is an irreversible neurodegenerative disease that results in the loss of mental function associated with the deterioration of brain tissue. It is the most common cause of dementia among people over the age of 65, affecting an estimated 5.5 million Americans.

Objective

- Predict cognitive decline for patients with normal, mild, moderate, and severe cognitive impairment
- Assess whether quantitative methods including PET standardized uptake value ratio and MRI volumetric analysis can be used to better predict possible dementia
- Identify clinical trial tests that are better or worse predictive indicators of dementia

USC Stevens Neuroimaging and Informatics Institute

Data Source

Alzheimer's Disease Neuroimaging Initiative (ADNI) public data designed to develop clinical, imaging, genetic, and biochemical biomarkers for early detection and tracking of Alzheimer's disease.

141 studies	**142,455** users
83,252 subjects	**161** countries

Target Variable

The Mini Mental State Examination (MMSE) is a 30-point questionnaire that is used extensively in clinical and research settings to measure cognitive impairment. It is commonly used in medicine and allied health to screen for dementia.

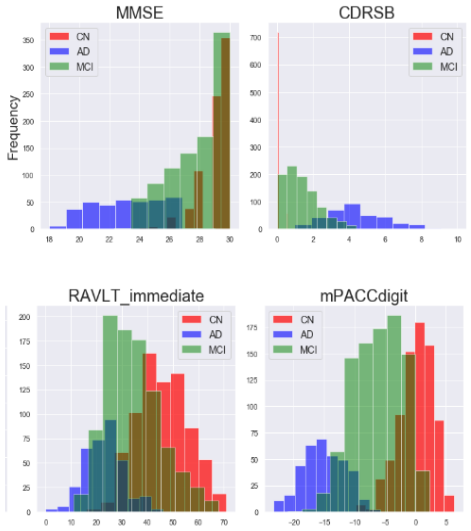
- **>=25: Normal**
- **20-24: Suggests mild dementia**
- **13-20: Suggests moderate dementia**
- **<=12: Suggests severe dementia**

Data Variable Definition

Variable	Description	Value Range	Missing at Baseline (%)
Primary Psychometric exams given:			
CDRSB	Clinical Dementia Rating: measures dementia progression, especially with early to mid level cognitive impairment	0-18	0
ADAS11/ADAS13	Alzheimer's Disease Assessment Scale: Measures cognitive and non-cognitive symptoms of AD	0-70	0.5/1.0
RAVLT	Rey Auditory Verbal Learning Test: measures auditory learning and memory	0-75	0.5
DX	Measures cognitive processes affected by dyslexia	0-50	0.5
MMSE (Target Variable)	Mini Mental State Examination: Widely used to assess cognitive function.	0-30	0

Variable	Description	Value Range	Missing at Baseline (%)
Primary PET measure			
TAU	Tau protein levels. Higher levels are associated with presence of AD	0-1	41.2%

Variable	Description	Value Range (such that 800-1000 considered decreased, 1000 considered no change, >1000 increased)	Missing at Baseline (%)
Primary MRI measure			
Hippocampus	Atrophy change analysis within the hippocampus	800-1100	28.9
WholeBrain	Atrophy change analysis within the Whole Brain	800-1100	19.2
Entorhinal Cortex	Atrophy change analysis within the Entorhinal Cortex	800-1100	29.9



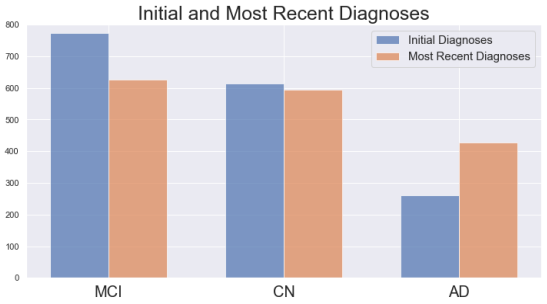


Image Source Citation: Introduction | Alzheimer's Project (ac209a-alzheimersproject.github.io)

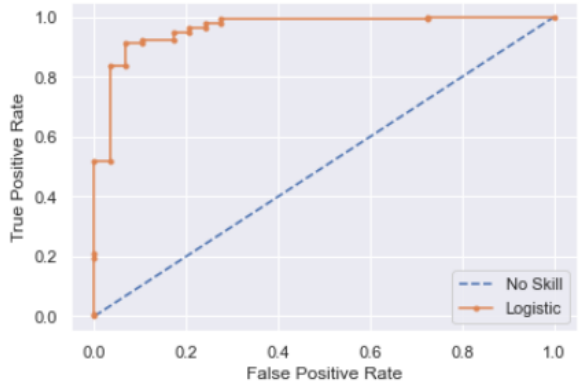
Results

Model Type	Precision	Recall	Accuracy
Clinical Data	85.11%	89.87%	87.89%
MRI + PET Imaging Data	72.12%	78.12%	74.78%
Clinical + MRI + PET Data	**92.20%**	**94.89%**	**91.15%**

Area Under the Curve (AUC)

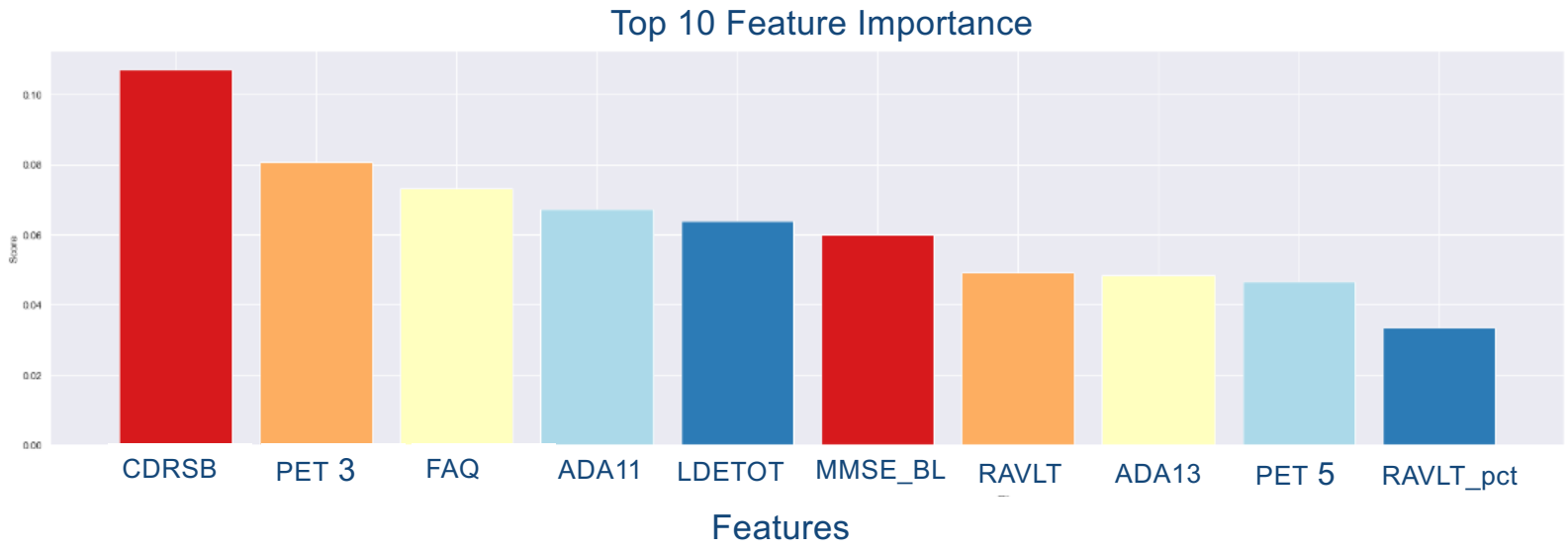
Estimates the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance.

```
No Skill: ROC AUC=0.500
Logistic: ROC AUC=0.962
```

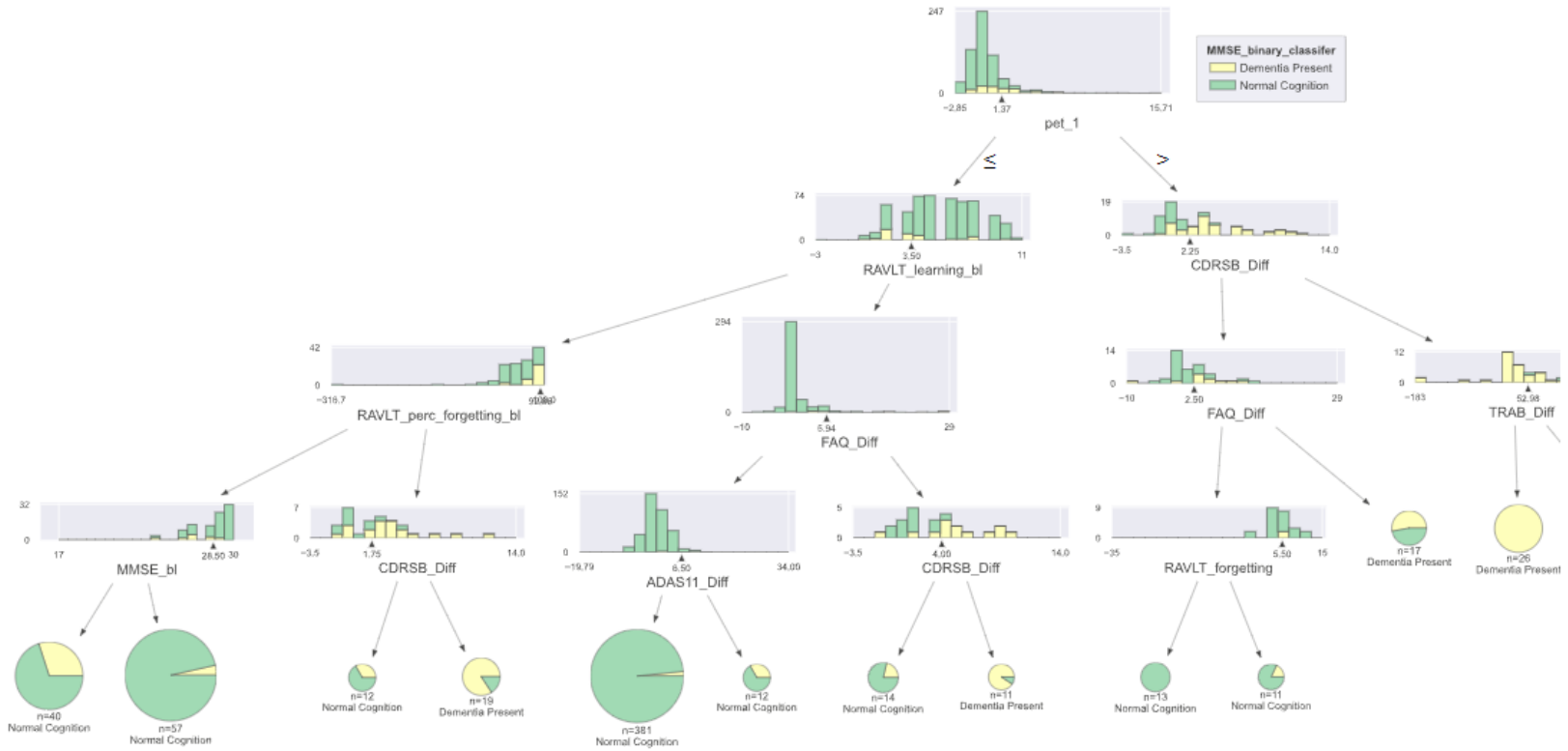


Feature Importance

Determines the degree of usefulness of a specific variable for the model and prediction

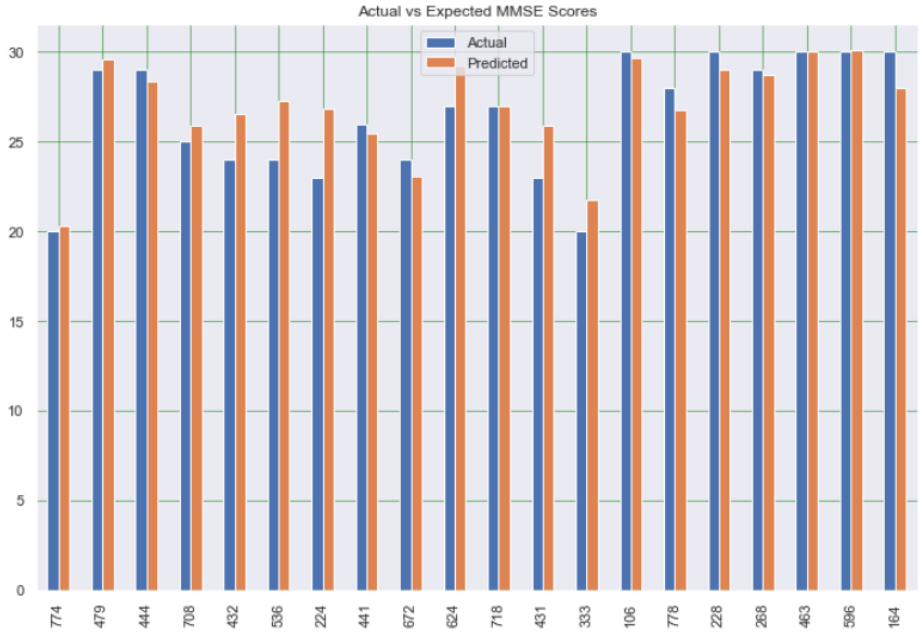


Model Decision Making



Actual Versus Expected

Model Type	Mean Squared Error	R^2
Clinical Data	2.44	0.85
MRI + PET Imaging Data	6.15	0.67
Clinical + MRI + PET Data	**1.91**	**0.91**

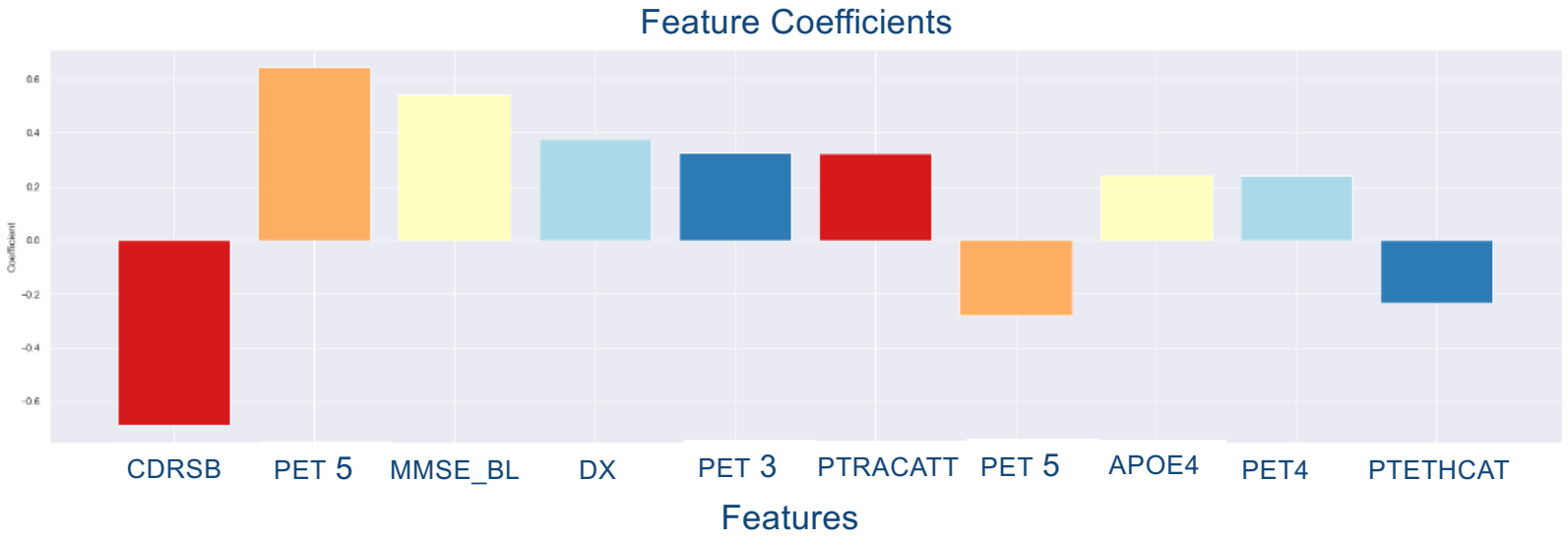


Predictive Formula to Assess Dementia

MMSE predictor formula = **μ** CDRSB (x) + **μ PET_5** (x) + **μ** MMSE_bl + (x) + **μ** DX + (x) **μ** PET_3 (x) ...

CDRSB: Measures Dementia Progress
PET_5: PET reduction composite
MMSE_bl: MMSE at Baseline

DX: Measures cognitive processes affected by dyslexia.
PET_3: Pet reduction composite



Thank you

Questions?