

Building Rome in a few months: How open-source technologies empower small pharma

Chendi Liao & Kasra Yousefi @ PHUSE US Connect 2025
March 17, 2025



Recursion

Techbio - tech-forward approach to **industrialize drug discovery**

Growing Clinical Capabilities

- Vision to build tech-driven clinical development
- Young and small in clinical-stage
- Current regulatory work done by CROs



Chendi Liao

DIRECTOR OF
STATISTICAL SCIENCE



Kasra Yousefi

DIRECTOR OF
BIOSTATISTICS

Limitations with External Vendor / CRO



High cost - typical Phase I trial cost \$3 - 5 million



Proprietary software / solutions



Lack of flexibility

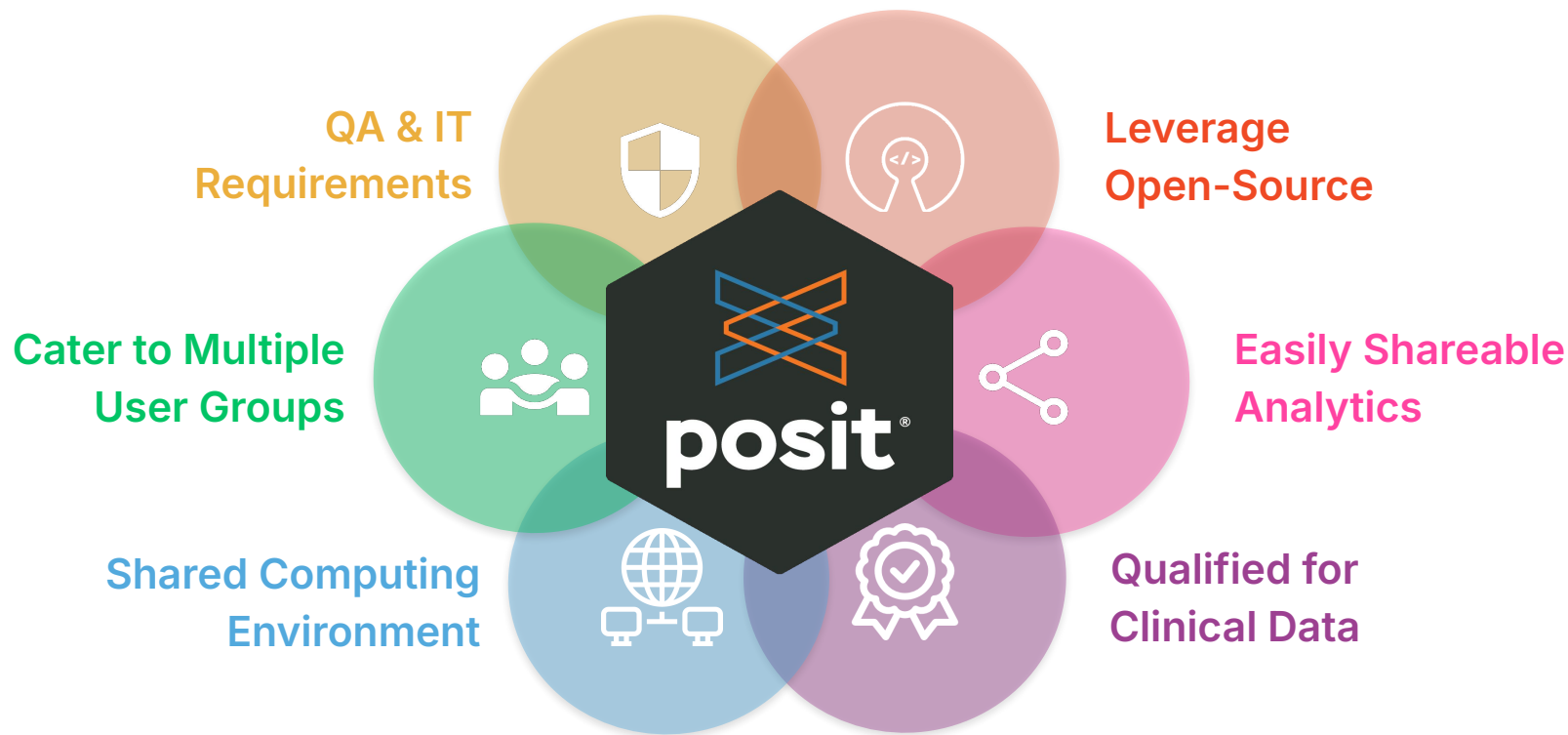


Lack of expertise in advanced analytics and visualizations



Long time between data and insights

Desired System for In-House Clinical Data Analytics





POSIT, OKAY

**YEAH!
LET'S GO!!**

...RIGHT?

Getting Resources in Place



Proposal, budget, demo/PoC



Management & stakeholder buy-in



Finding expertise

- Experience with GxP data
- Establishing user stories and use-cases



Securing IT bandwidth - Infrastructure, Security, Engineering

**TWO WEEKS
LATER...**



Accelerated insights and decision-making



REC-XXX-123 Clinical Data Review Demo Report

Recursion Pharmaceuticals Confidential Document - For Internal Use Only.
Disclaimer - This demo/examples is created with dummy data

Prepared by Chendi Liao

Data Snapshot Date: 2024-XX-XX; Demo report generated on: 2024-08-20

Individual Patient Profile

Subject ABC-07

Demographics and Baseline

ABC-07: Demographics and Baseline						
Demographics			Baseline			
Age/Sex/Race/Ethnic ¹	Height at Baseline [cm]	Weight [kg]	BM [kg/m ²]	Disease K-score	Status	
70F/Asian/NH	149	47.8	21.5	70	NA	
¹ NH = Not Hispanic or Latino, H = Hispanic or Latino						

Disease Characteristics

ABC-07: Disease Characteristics		
Disease History		Prior Therapy Procedures
Duration of Disease (yrs)	Age at Symptom Start	Age at Diagnosis
1	61	69
Disease Grade	Radiation	Chemotherapy
Grade >=2	>1	N
Surgery	Number of Prior Radiotherapies	
>1	>1	

Treatment and Disposition Overview

ABC-07: Overview of Treatment and Disposition Info		
Treatment Group Info	Disposition Status	Exposure Info
Stable		

Clinical Data Exploration Demo App

Biometrics @ Recursion

App Info Data Table Variable Browser **Demographic Table** Forest Plots Kaplan Meier Plot Response Table Time to Event Table Cross Table Cox Reg Logistic Reg MMRM

Binary Response ANCOVA Report previewer

Reporter

Encodings
Dataset: ADSL

Select Column Variable(s)
Select
[L] ARM Description of Planned [+]

☒ Add All Patients column

Summaries Variables
Select
[L] SEX [H] ETH [NIC] ETH

> Additional table settings

Show Warnings Show R code

	A: Drug X (N=32)	B: Placebo (N=26)	C: Combination (N=22)	All Patients (N=80)
Sex				
n	32	26	22	80
F	17 (53.1%)	12 (46.2%)	11 (50%)	40 (50%)
M	15 (46.9%)	14 (53.8%)	11 (50%)	40 (50%)
Ethnicity				
n	32	26	22	80
HISPANIC OR LATINO	2 (6.2%)	0	4 (18.2%)	6 (7.5%)
NOT HISPANIC OR LATINO	26 (81.2%)	21 (80.8%)	13 (59.1%)	60 (75%)
NOT REPORTED	2 (6.2%)	4 (15.4%)	3 (13.6%)	9 (11.2%)
UNKNOWN	2 (6.2%)	1 (3.8%)	2 (9.1%)	5 (6.2%)
Age Group				
n	32	26	22	80
<45	29 (90.6%)	26 (100%)	21 (95.5%)	76 (95%)
>=45	3 (9.4%)	0	1 (4.5%)	4 (5%)

n represents the number of unique subject IDs such that the variable has non-NA values.

Active Filter Summary

Data Name	Obs	Subjects
ADSL	80/80	80/80

Active Filter Variables

ADSL

SEX
F, M
☒ F (40/40)
☒ M (40/40)

Add Filter Variables

Add ADSL filter
Select variable to filter

Disclaimer: All demo/examples are created with dummy data

Accelerated insights and decision-making (Clinical Trial Operations)

EnrollSight

Introduction

Trial Focus

Enrollment Analytics

Predictive Analytics

Focus Your Trial Search

Condition/Disease
endometrial

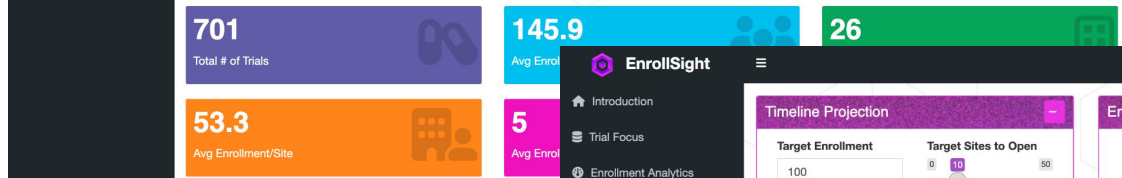
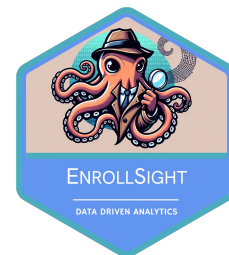
Eligibility Age
☐ Child (birth - 17)
☐ Adult (18 - 64)
☐ Older adult (65+)

Study Phase
☐ Early Phase 1
☐ Phase 1
☐ Phase 2
☐ Phase 3
☐ Phase 4
☐ Not applicable

Study Start
From
To

Other Terms
Terms to narrow your search, eg. drug name, genetic m

Search



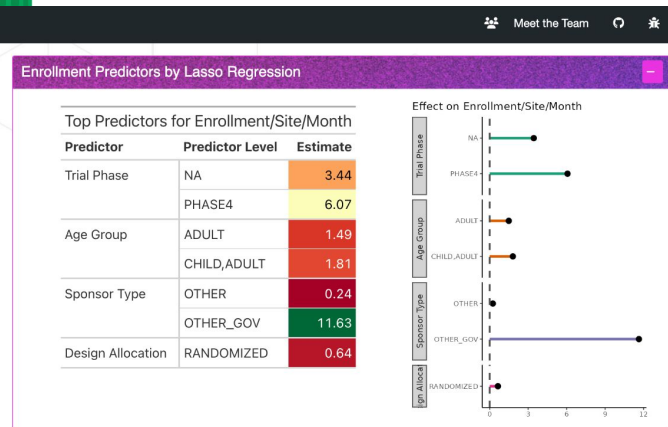
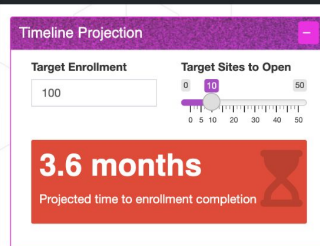
EnrollSight

Introduction

Trial Focus

Enrollment Analytics

Predictive Analytics



Lessons Learned



Start early!
Don't wait



Leverage Posit
support



Create PoC/MVP
to engage



User experience
matters!



Establish roadmap
& timeline



Do consider
consulting services

Plans in Motion



**GxP Data Lake
Connecting to Posit**



Validated SCE

Where We Were



Where We Are Now



Where We Want to Be

Vendor-dependent

(Semi) Self-sufficient

Self-sufficient with fully validated SCE

Costly proprietary software

Free open-source tools

Connection to GxP Data Lake

Wait time in days/weeks

Accelerated insights

Real-time insights

Biometrics Team



Patients



Clinical Development Team

Decode Biology to Radically Improve Lives



Recursion