

Transforming Study-Specific Reporting Trends

Date: **17 Nov 2025**



Authors: Nayak Archana; Ophoff, Jill; Van Yper, Stefan; Walsh, Simon; Reddy, Bhargava, Johnson & Johnson R&D

PHUSE workstream: Analytics & Statistics (AS) AS02

Presenting author:
Archana Nayak, PhD

Advanced Insights and solutions

Data Science and Digital Health - Global Functions www.jnj.com



Disclaimer

This presentation is for informational purposes only and does not represent professional guidance or advice.

Any views and opinions expressed during this presentation are those of the presenters and do not necessarily reflect the views or policies of J&J Innovative medicine Research & Development, LLC, or any other company in the Johnson & Johnson Family of Companies. • © J&J Innovative medicine Research & Development, LLC. This presentation may not be reproduced, modified, or copied, in whole or in part, without the express written permission of J&J Innovative medicine Research & Development, LLC.

Importance of Risk-Based Quality Management !!



INTERNATIONAL COUNCIL FOR HARMONISATION OF TECHNICAL
REQUIREMENTS FOR PHARMACEUTICALS FOR HUMAN USE

ICH HARMONISED GUIDELINE
GUIDELINE FOR GOOD CLINICAL PRACTICE
E6(R3)

Final version
Adopted on 06 January 2025

ICH E6(R3) Guideline

3.10.1 Risk Management

A proportionate approach to the identification and management of risk is described below:

3.10.1.1 Risk Identification

The sponsor should identify risks that may have a meaningful impact on critical to quality factors prior to trial initiation and throughout trial conduct. Risks should be considered across the processes and systems, including computerised systems, used in the clinical trial (e.g., trial design, participant selection, informed consent process, randomisation, blinding, investigational product administration, data handling and service provider activities).

3.10.1.2 Risk Evaluation

The sponsor should evaluate identified risks and existing controls in place to mitigate the risk by considering:

- (a) The likelihood of harm/hazard occurring;
- (b) The extent to which such harm/hazard would be detectable;
- (c) The impact of such harm/hazard on trial participant protection and the reliability of trial results.

3.10.1.3 Risk Control

Risk control should be proportionate to the importance of the risk to participants' rights, safety and well-being and the reliability of trial results. Risk mitigation activities may be incorporated, for example, in protocol design and implementation, monitoring plans, agreements between parties defining roles and responsibilities, and training.

DATA4YOU

One system for all

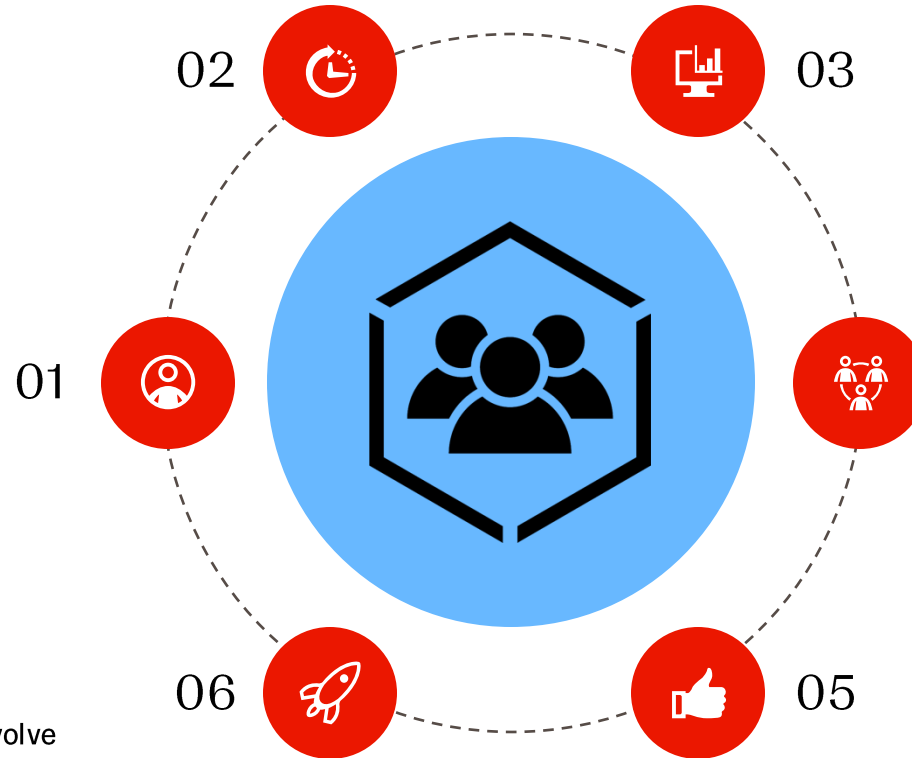
Near Real Time Data

Intuitive and User-Friendly

Good Performance, Easy navigation, multi-window and multi-monitor view, easy sorting and filtering, user customization, patient profiles

Future Ready

AI/ML capabilities, New Standards, Let's Evolve



Enriched Data

SDTM, DRM, IWRS, CTMS, Coding, etc.

Cross-Functional Collaboration

Review status by role (DM, CM, TA), tasks by role, inquiries and comments

Good Quality Visualizations



AI/ML support to benefit multiple roles in DATA4YOU

Data Management (DM)

01



Clinical Programming (CP)

02



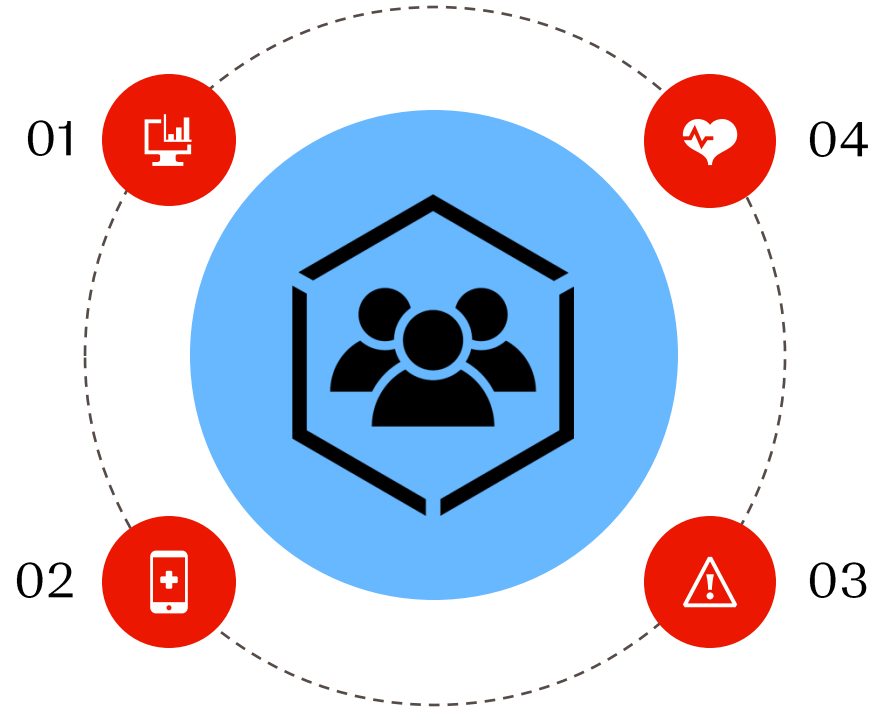
Medical Review Group (MR)

04



**Risk Management
Central Monitoring (CM)**

03



AI in clinical data management & central monitoring

AI/ML Approaches – even prior to GenAI



Central Statistical Surveillance (CSS)

- Detects unusual patterns, potential data quality issues at investigative sites
- AI/ML to conduct holistic data examination, from numerical (e.g. vital signs, Tumor assessments) to categorical (e.g. Adverse events, dispositions)
- Central monitors reporting overall reduction in review time by 45%



Site Profile Tool (SPT)

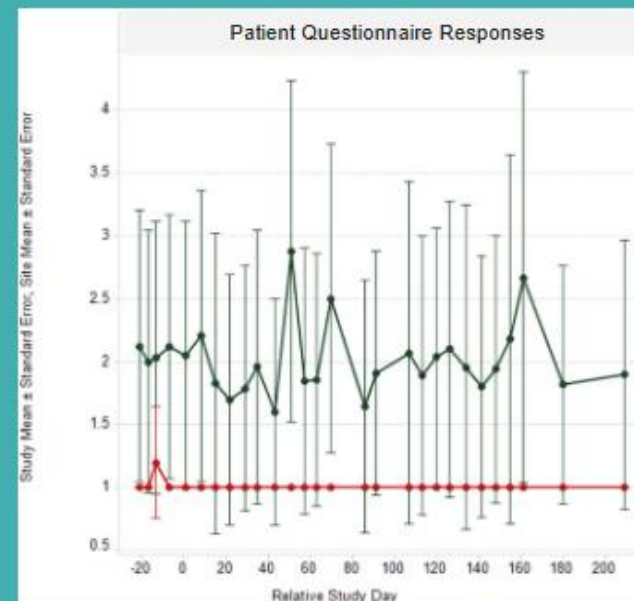
- Identifies areas of potential risk across business with in-depth review of KRIs (Key risk indicators)
- Creates a composite score for each investigative site with R-A-G presentation
- PD classification



AE/LAB Discrepancies (ALIGN)

- Automatically identify grade and dates discrepancies for lab related adverse events
- Implemented AI engine in-house, saving ~300K per year

Results at site in question are more alike compared to the study norm:



LEGEND:

- **Green Line and Error Bars:** patient assessment results across all sites in study
- **Red Line and Error Bars:** site in question showing results differing from study norm and with almost no variation/error bars

Study Specific Reports (SSR) is key oversight Tool for Central Teams



Study Specific Reports (SSR) are determined as part of Risk-Based Monitoring (RBM) process and through protocol de-risking. (ICH3)



SSR KRIs should have a direct link with Critical Data Review and enlist Major Protocol Deviations.

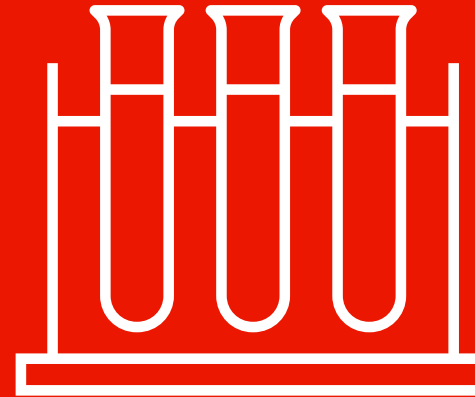
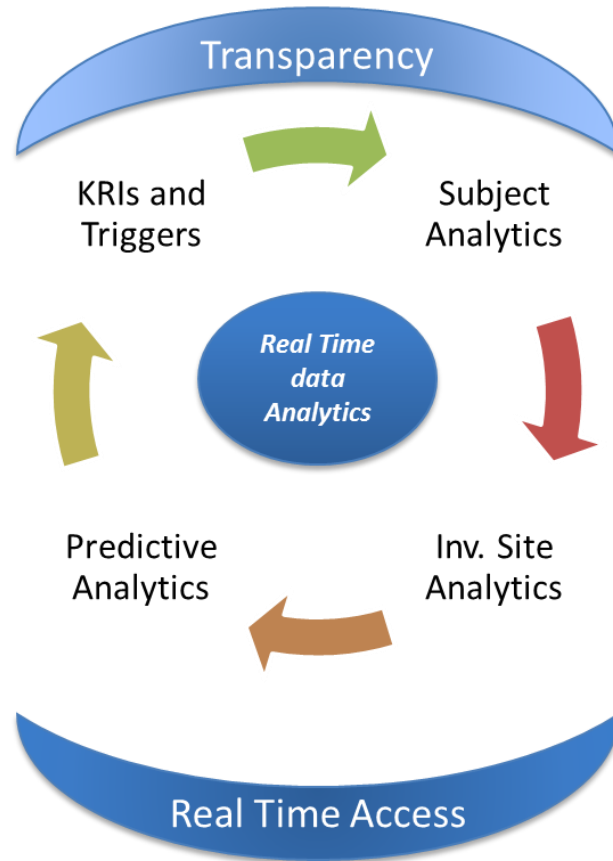


SSRs highlight programmable inconsistencies that draw data from different sources including IV/WRS, EDC, ePRO, Biospecimen data, etc...



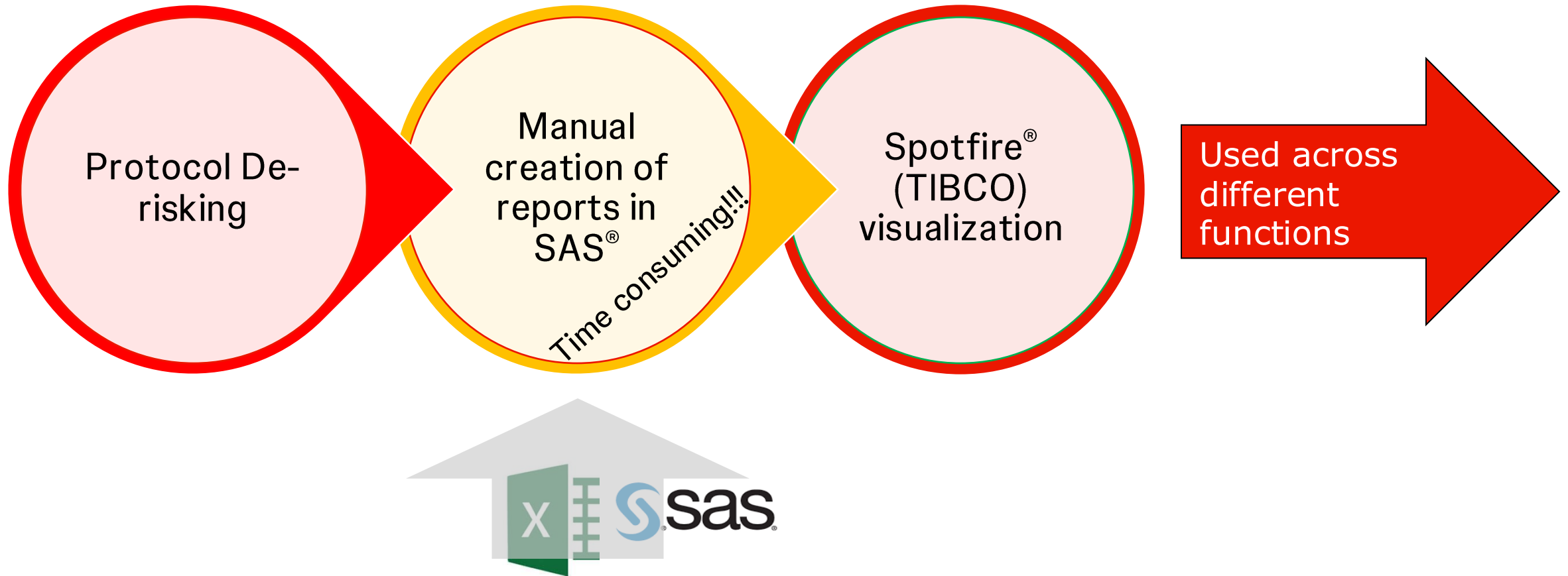
The insights are actionable in D4U platform and shared across functions responsible for monitoring the trial.

Focus on SSR BE trends



Study-Specific Report
Biospecimen samples
Trends

Previous SSR process



Opportunity Identification

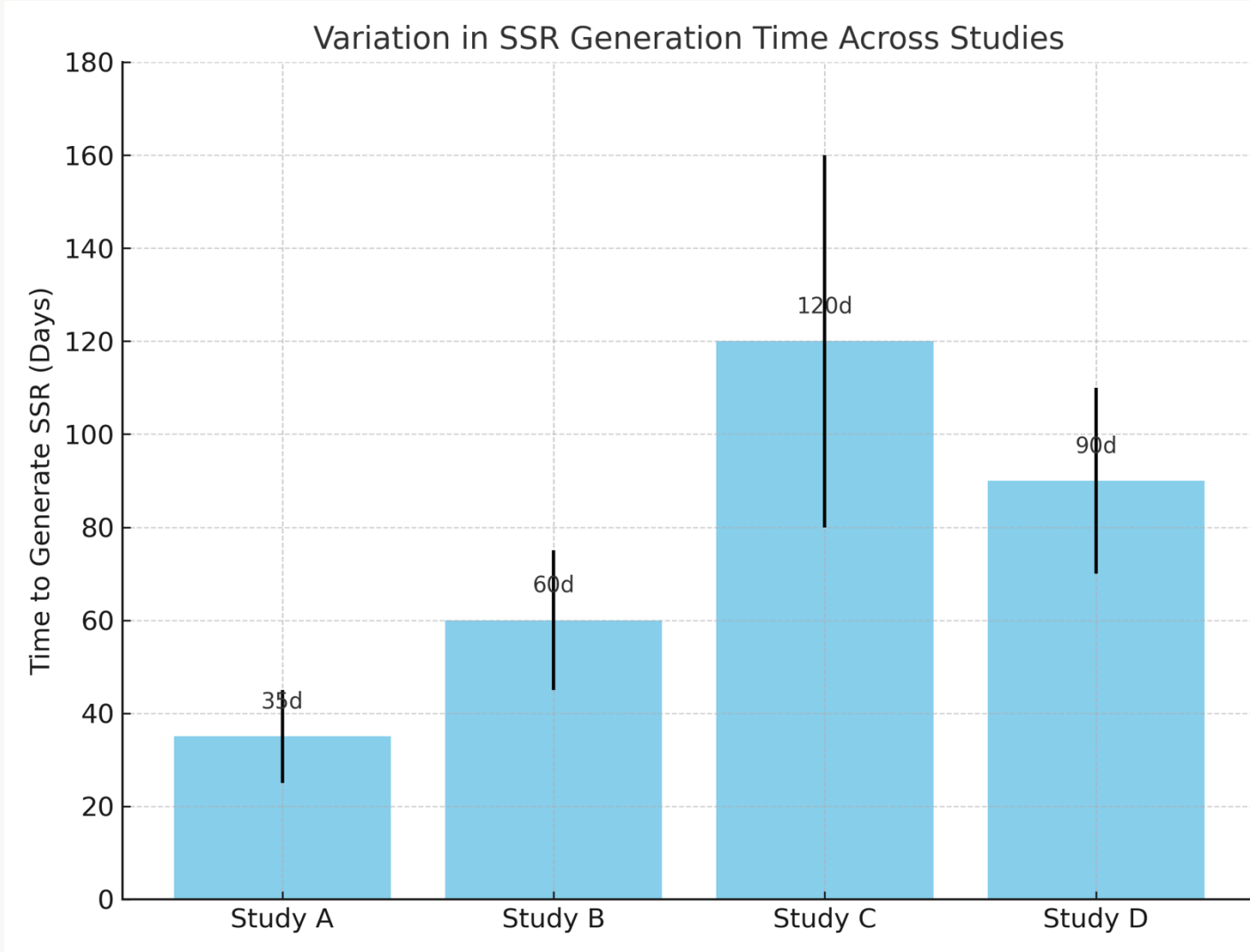
Problem Statement & Data Science Proposal

Problem Statement:

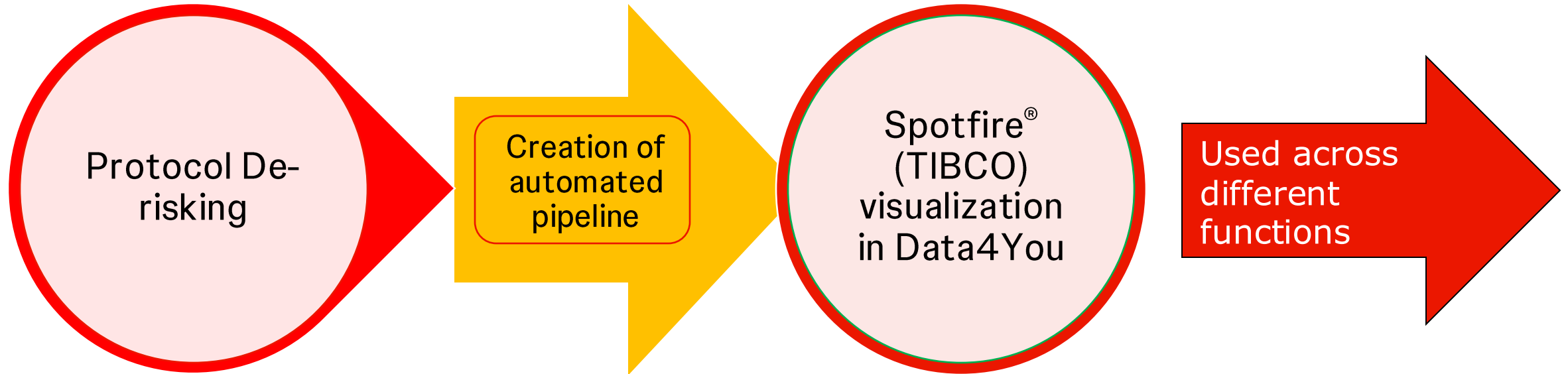
- Large variation in time needed for creating SSR reports
- Very dependent on trial specific Protocol's SOA(Schedule of Assessments) non scalable
- Line by line data reconciliation

Data Science Idea:

- Can we leverage data driven way of programming SSRs?

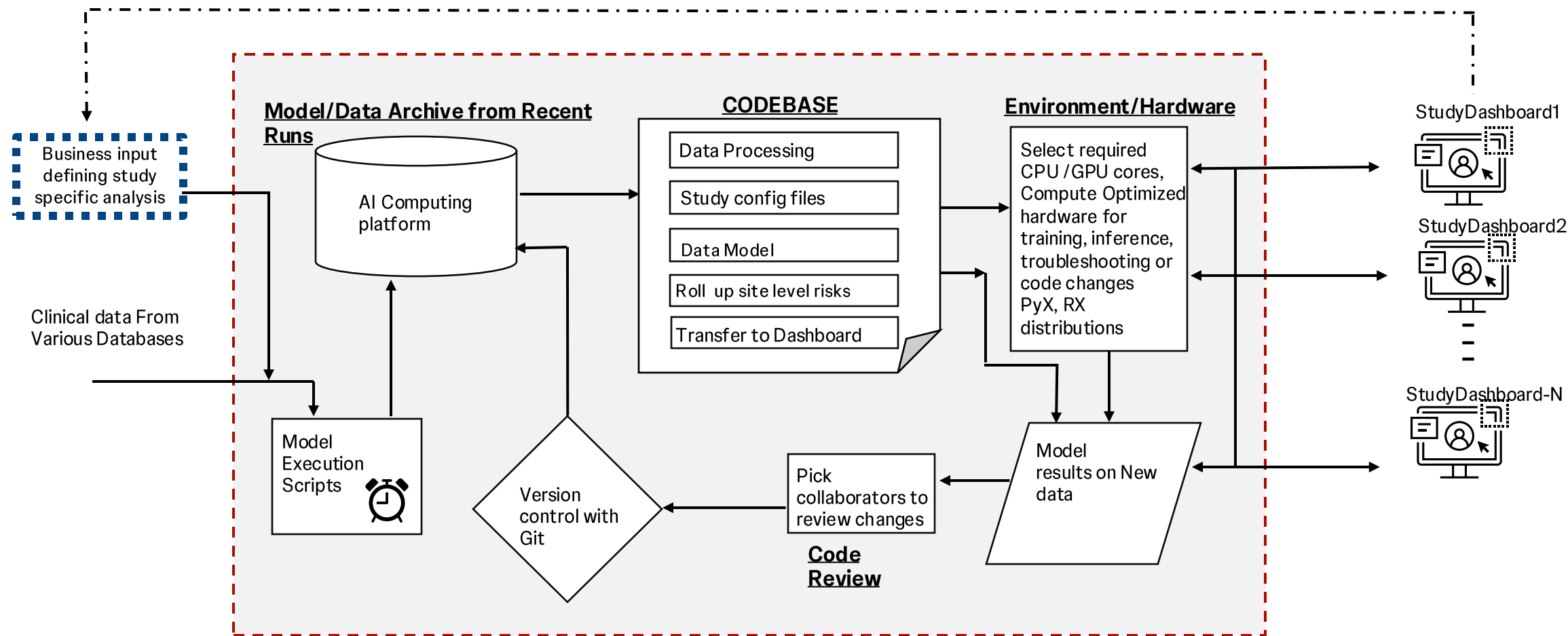


Current enhanced SSR Process



Data architecture

Automated Study data specific pipeline



Extraction of meta data from input files

- Raw Data Ingestion (EDC, Labs, Vendors)
- Config-Driven Transformation (YAML/JSON)
- Report Generation & Visualization
- Supports batch/streaming
- Decouples study specific logic from overall codebase

Variable names	Missing values	Present values	Unique values
STUDYID	0	936	1
RDOMAIN	0	936	1
USUBJID	0	936	486
IDVAR	0	936	1
IDVARVAL	0	936	16
QNAM	0	936	1
QLABEL	0	936	1
QVAL	0	936	2
QORIG	0	936	1
QEVAL	0	936	1

Domain variables	Missing values	Present values	Unique values
raw file	/mnt/STUDY_ID/data/SV	/mnt/STUDY_ID/data/SV	/mnt/STUDY_ID/data/SV
STUDYID	0	54593	1
SVTESTCD	0	54593	95
SVTEST	0	54593	95
SVCAT	0	54593	4
SVSTRESC	0	54593	2660
SVSTRESN	7604	47535	2539
SVSTRESU	0	54593	10
SVSTNRLO	6603	48536	44

Study specific configuration

Automated set up due to modular approach

- Modular, reusable components
- Code config templates
- Harmonized data formatting
- Near real-time updates

TEST	VISIT	DACAT	OSCAT	TRCAT	LBCAT
DISPSTAT	50	3			
ECOG101	83		1		
DIAMETER	7			1	
BILDIR	99				1
BILI	105				2
GGT	104				1
GLUC	105				2
HCAB	7				1
HCG	4				2
RBC	104				2
HIVAB	6				1

Test Name	VISIT	MI counts	SR counts	LB counts	Criteria
ULCER	8	5			INCLUDED
INDURDIA	1		5		EXCLUDED
INTP	2		0		INCLUDED
BASO	36			29	INCLUDED
BASOIM	1			0	EXCLUDED
BASOIMLE	1			0	EXCLUDED
BASOLE	36			31	INCLUDED

LBTEST CD	LBCAT	LBSPEC	LBSTRESU
RBC	URINALYSIS	URINE	/HPF
RBC	HEMATOLOGY	BLOOD	10^12/L
BILI	CHEMISTRY	BLOOD	umol/L
BILI	URINALYSIS	URINE	-

Data driven sampling technique

Automated extraction of sample collection timepoints

median per GROUPINGVARIABLE, VISIT and BESPPURP, SAMPLE_COLLECTION_TIMEPOINT

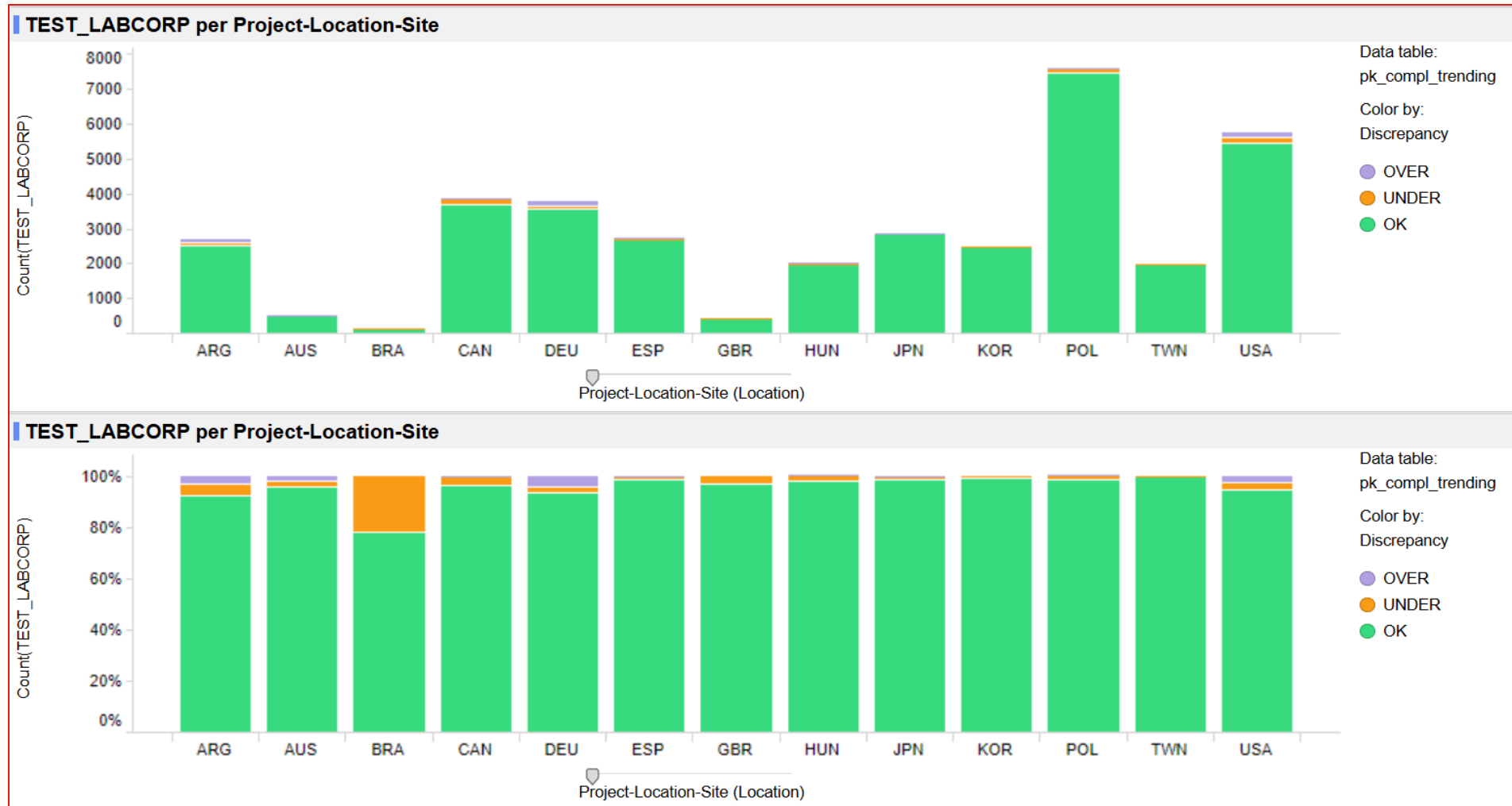
(Column Names) » BESPPURP » SAMPLE_COLLECTION_TIMEPOINT

GROUPINGVARIABLE	VISIT	median							
		SM01/PLASMA BIOMARKER		SM02/SERUM IMMUNOGENICITY		SM03/DNA	SM04/C SF	SM07/C SF BM STORAGE	
		NA	PREDOSE	NA	PREDOSE	NA	PREDOSE	PREDOSE	
GROUPINGVARIABLE » VISIT	WEEK 0 (DAY 1)	-	1.00	-	1.00	1.00	0.00	0.00	
	WEEK 2	1.00	-	1.00	-	-	-	-	
	WEEK 8	-	1.00	-	1.00	-	-	-	
	WEEK 10	1.00	-	1.00	-	-	-	-	
	WEEK 24	-	1.00	-	1.00	-	-	-	
	WEEK 26	1.00	-	1.00	-	-	-	-	
	END OF STUDY (WEEK 208)	1.00	-	1.00	-	-	-	-	

- Determined
- Undetermined

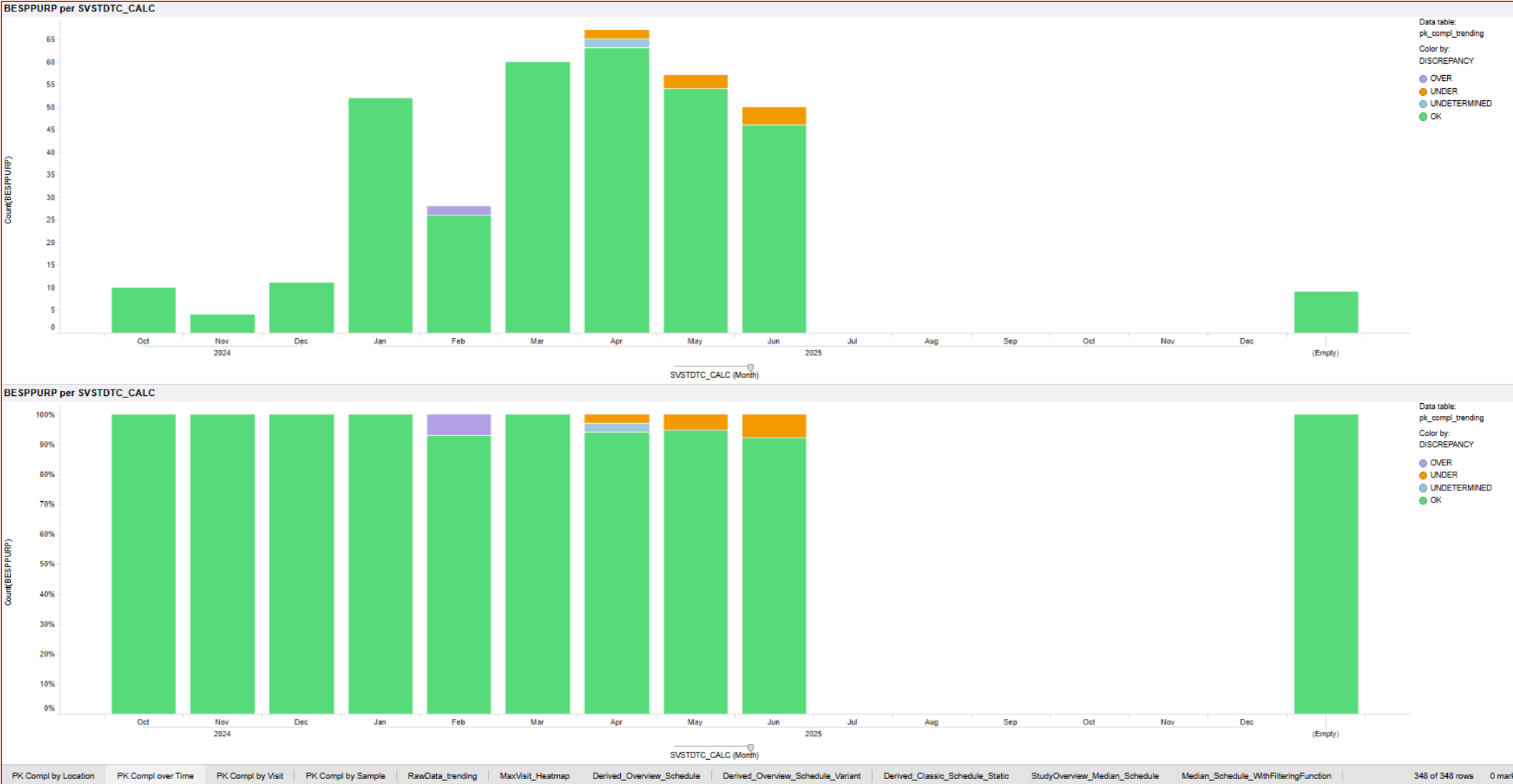
Key insights

By Location



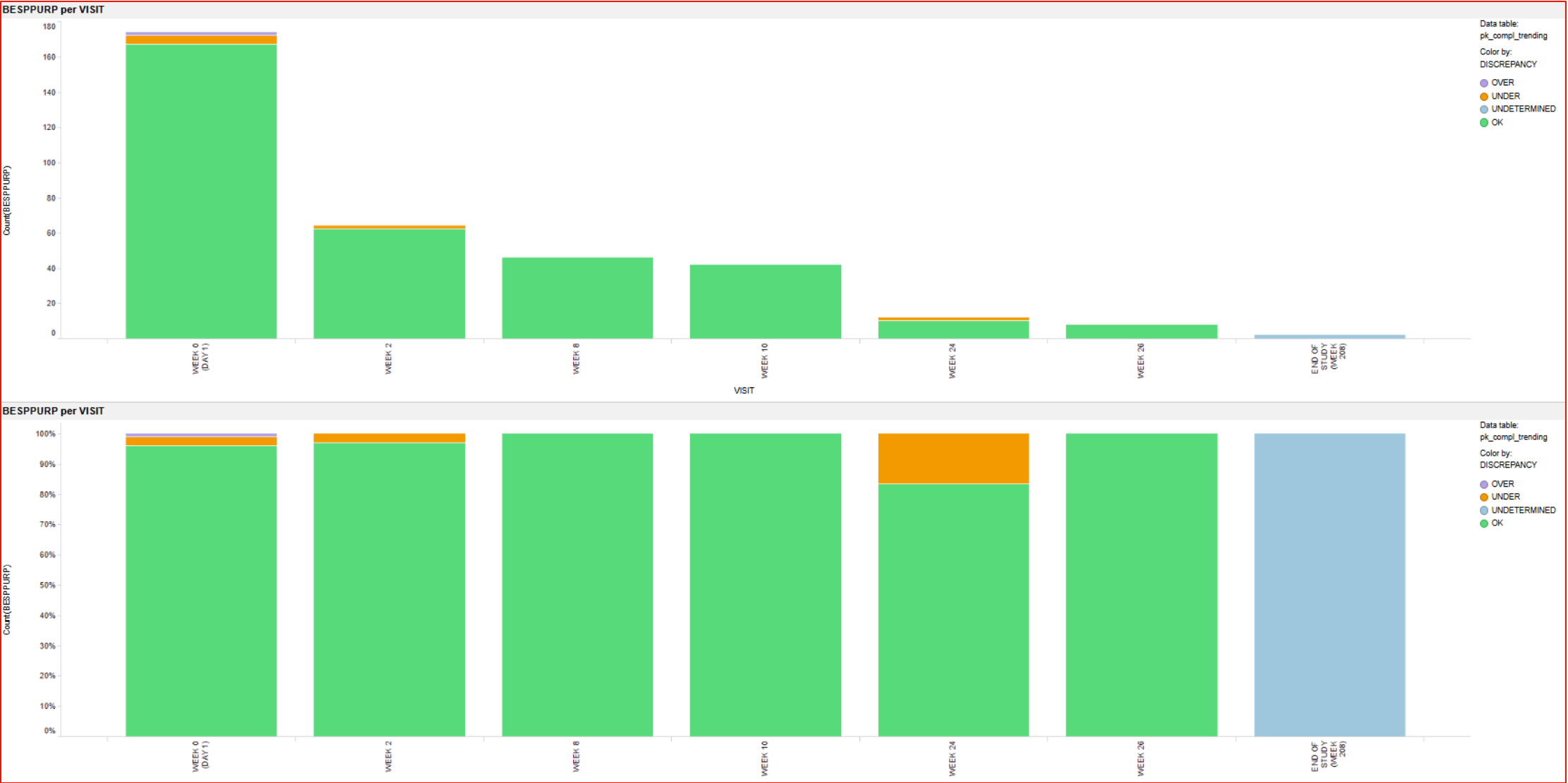
Key insights

Over Time, By Visit, By Sample



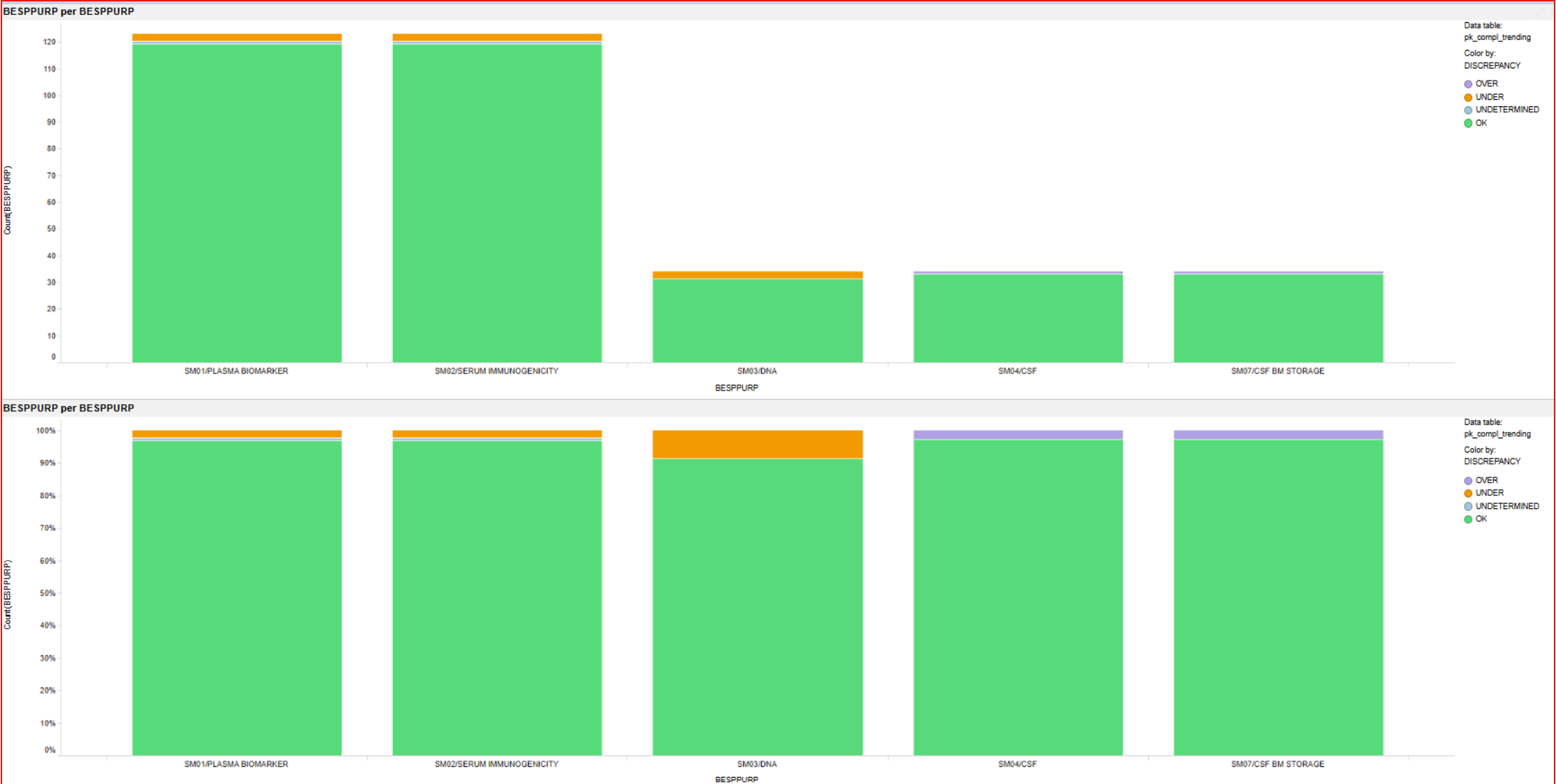
Key insights

By Visit



Key insights

By Sample



Future enhancements

- Adding possibility for user driven groupings to accommodate Protocol amendment updates
- Downstream data tables to be leveraged into agentic pipeline to derive site level insights



Summary

AI/ML approaches enhance project deliverables and user experience

 Summary	 Accuracy	 Efficiency
✓ Automation of study specific configuration files to streamline the process ✓ Lesser manual interventions ✓ Expansion to more domains	✓ Data Driven sampling schedule fastens study set-ups ✓ Study specific configuration file serves as reference for customized analysis too	✓ Reduce time to report generation ✓ Users to gain efficiency (speed of analysis completion) to insights



Acknowledgments

Many thanks to conference committee for this opportunity to present our work !

Many thanks to Stefan Van Yper; Jill Ophoff; Simon Walsh; Bhargava Reddy for contributions and continuous support in this project and *all team members of Data Science and Digital Health - Global Functions, DMCM teams, Johnson & Johnson Innovative Medicine R&D*