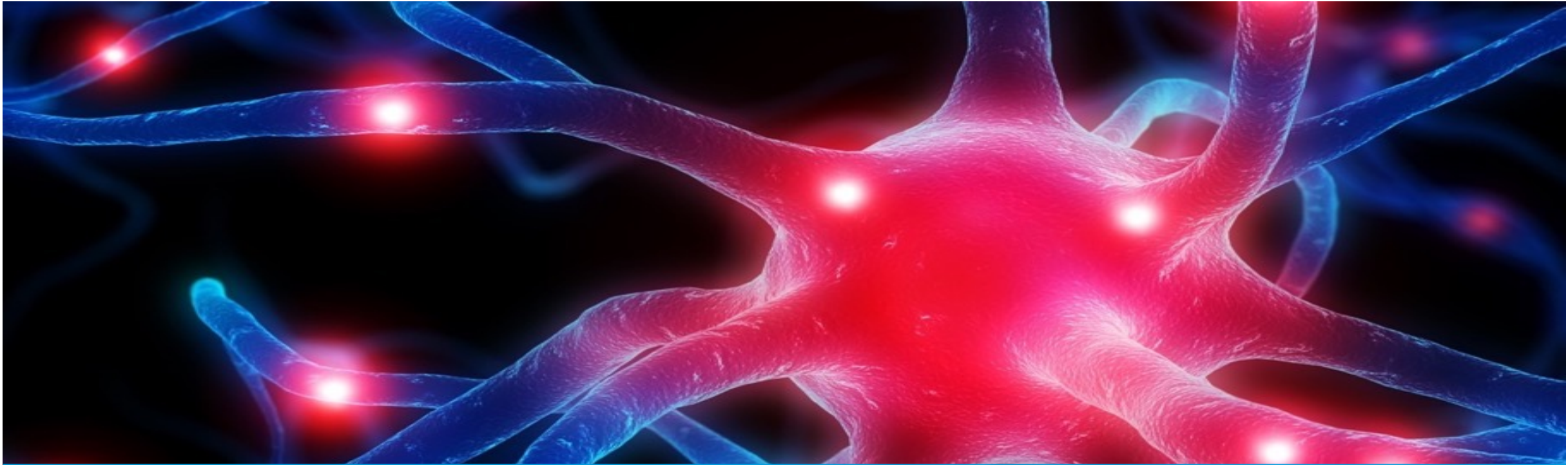




Advanced analytics for data quality assessment in countries and regions affected by crisis

7 Nov 2023

PHUSE EU Connect 2023, AR02

Presenting author 1: Archana Nayak, Janssen R & D, High Wycombe, United Kingdom, Email: anayak32@its.jnj.com

Presenting author 2: Md Umarfaruque, ZS, Janssen R & D, Email: umd@its.jnj.com

Active neuron

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 Janssen Research & Development, LLC, or any other company in the Johnson & Johnson Family of Companies.

© Janssen Research & Development, LLC. This presentation may not be reproduced, modified, or copied, in whole or in part, without the express written permission of Janssen Research & Development, LLC.

Agenda

Central Statistical Surveillance (CSS)



What?

 **180+** data processing cycles for  **100+** Janssen studies across  **multiple** therapeutic areas

Scope and outcomes



1000+ outliers identified amongst multiple sites

100+ potential findings identified as sites patterns for further investigation



100+ key insights discussed for designing corrective and preventive actions

100+ study teams informed with dynamic communications almost real time data insights

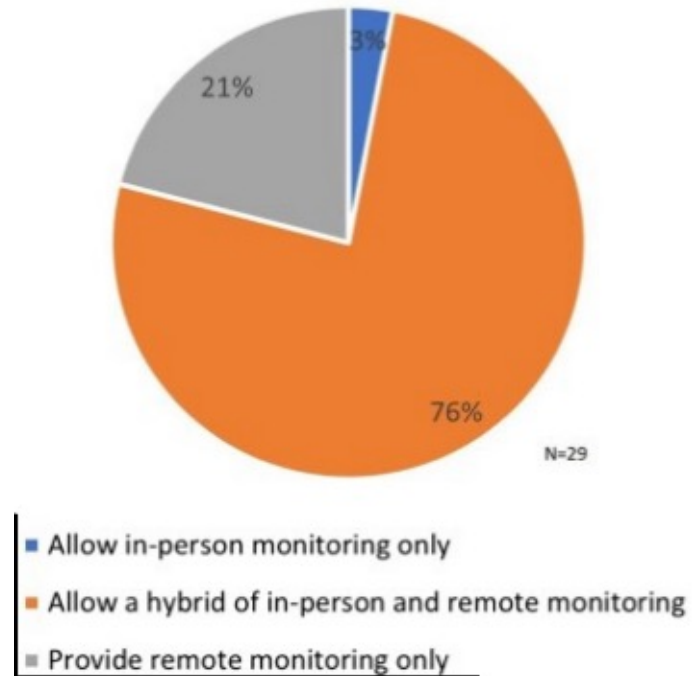


100+ potential findings translated into action



Need for country level comparisons in CSS

Issue



Graph credit: George et al, 2023 which illustrates an example about of impact of the COVID-19 pandemic and its quantitative impact on cancer center clinical trial operations highlighting the need for robust remote monitoring tools and capabilities

Aims to Uncover Irregularities in Country-Specific Data

Data quality

issues at country and/or region level

Data Pattern

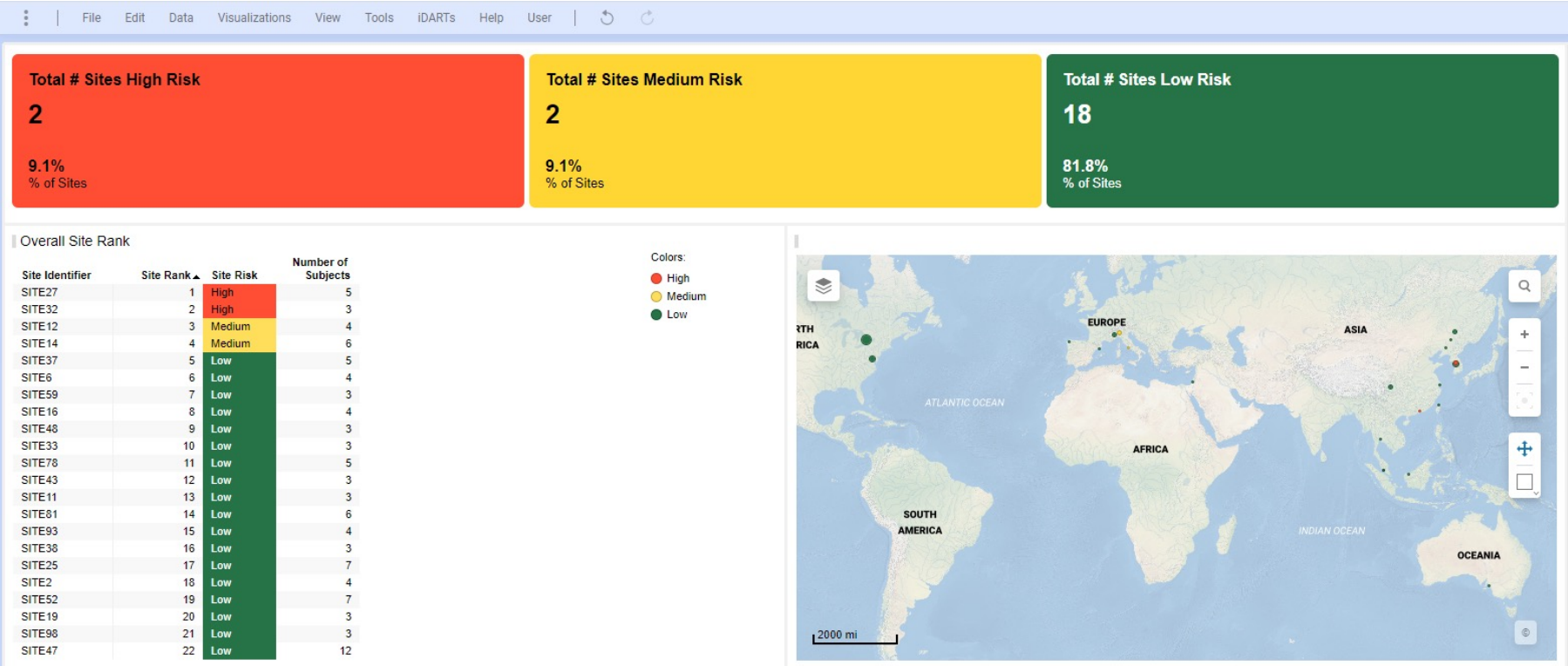
country-specific patterns, local practices, etc.

Local influence

impact of local and regional crisis

Systemic issues

country-level systemic issues including faulty batch of equipment, etc.



Solution: Roll up of site level outcomes of CSS to country level

Country Rank

Taking weighted average from both dimensions using Number of subjects and Site Risk score

Country Panel

Summary table to display total counts

Country signal Panel

Sites grouped together as lists with same signals with percentage

Country signal Details

Signal magnitude at test code level

Risk score of country

$$\varphi_i = \sum_{j \in \xi_i} \frac{\vartheta_j}{\left(\frac{\psi}{\eta}\right)} \chi_j$$

Equivalent site risk score

$$\chi_i = \begin{cases} 1 & \zeta_i = \text{High} \\ 0.5 & \zeta_i = \text{Medium} \\ 0 & \zeta_i = \text{Low} \end{cases}$$

Risk category of country

$$\varrho_i = \begin{cases} \text{High} & \varphi_i \geq 0.66 \\ \text{Medium} & 0.33 \leq \varphi_i < 0.66 \\ \text{Low} & \varphi_i < 0.33 \end{cases}$$

φ_i : Risk score of country i

ϑ_j : Total number of subjects on site j

ψ : Total number of subjects enrolled in study

η : Total number of country

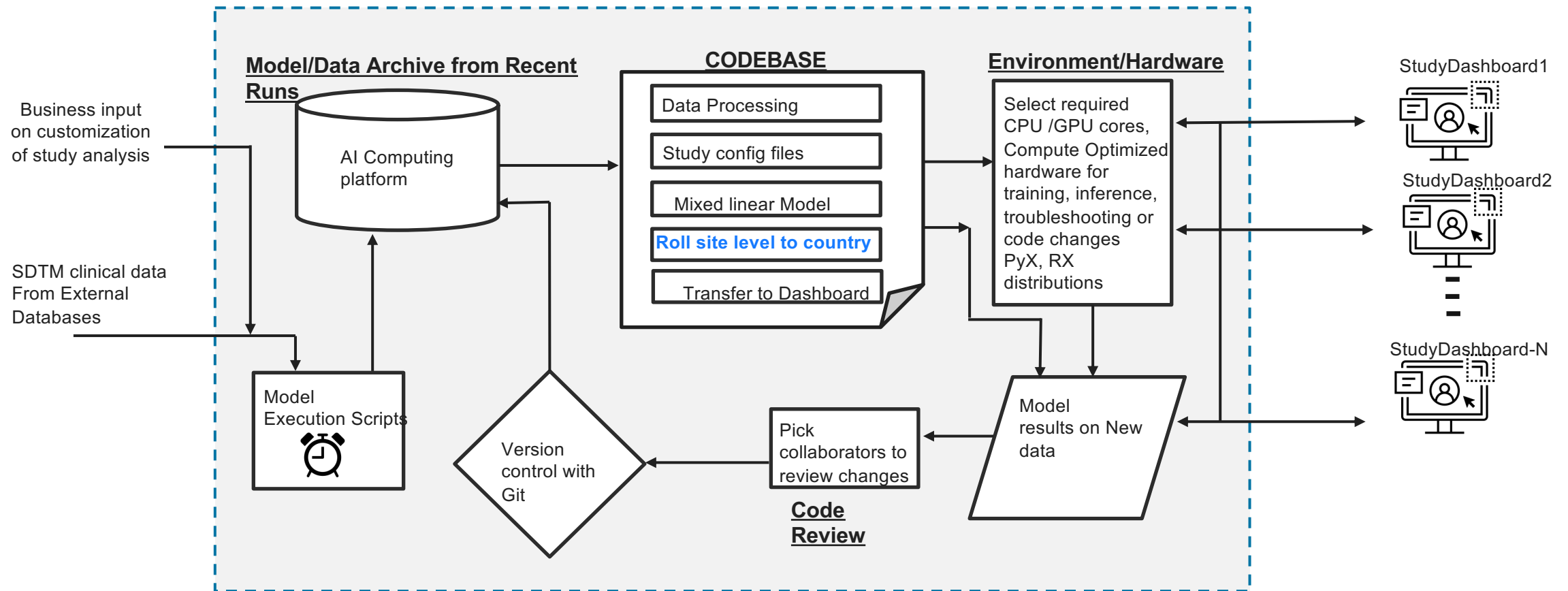
ξ_i : List of sites in country i

χ_i : Risk score of site j

ζ_i : Risk category of site i

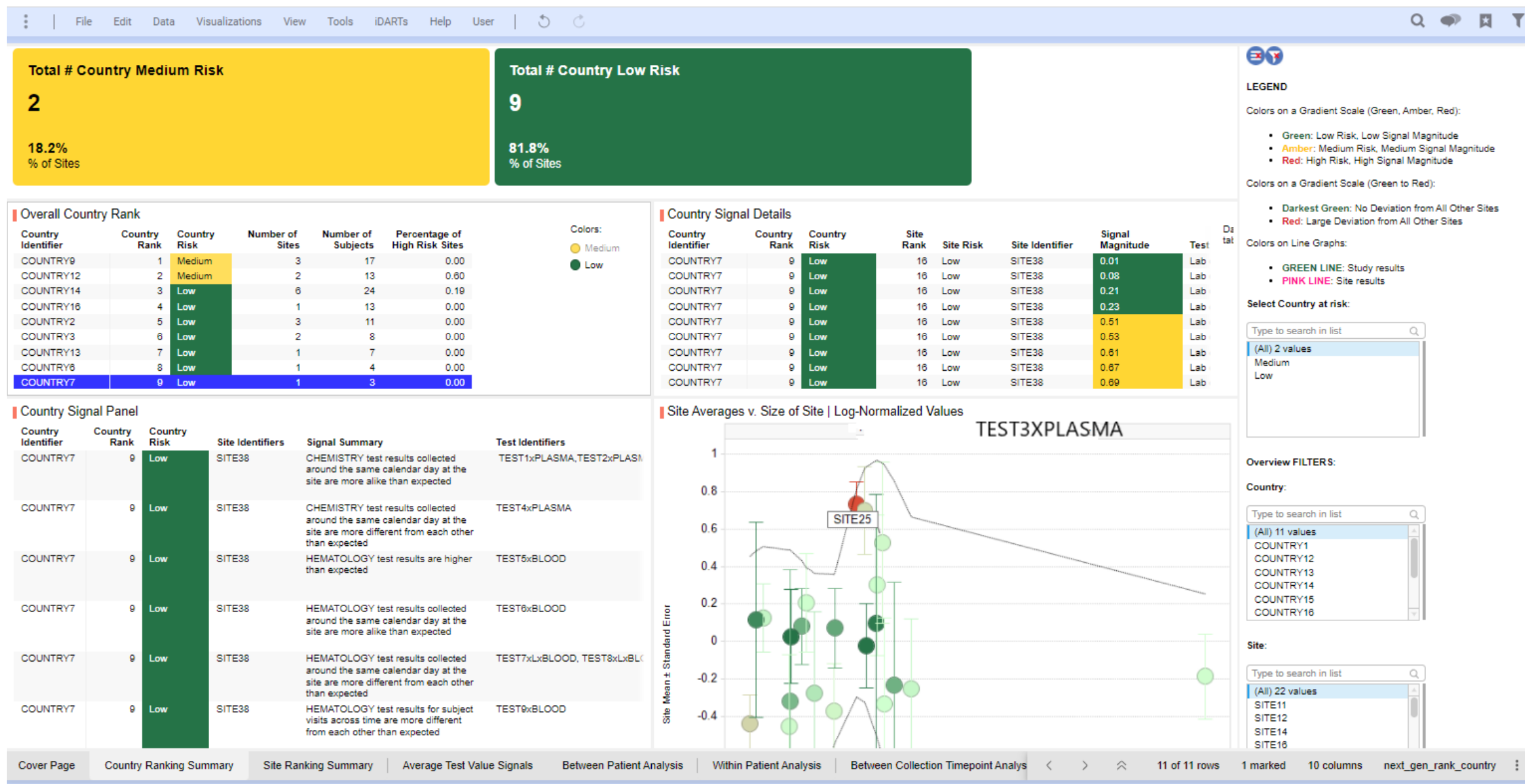
ϱ_i : Risk category of country i

Data pipeline we built



Country Review functionality

How

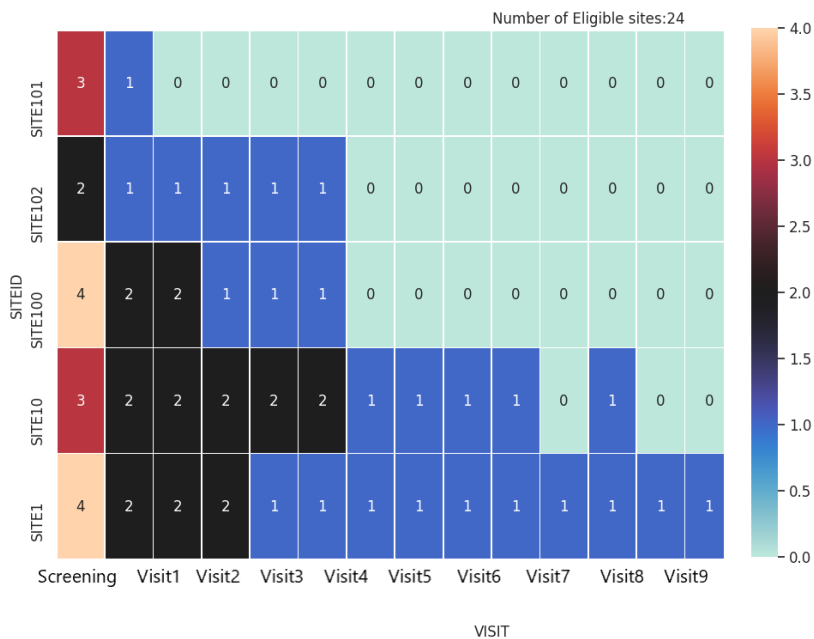
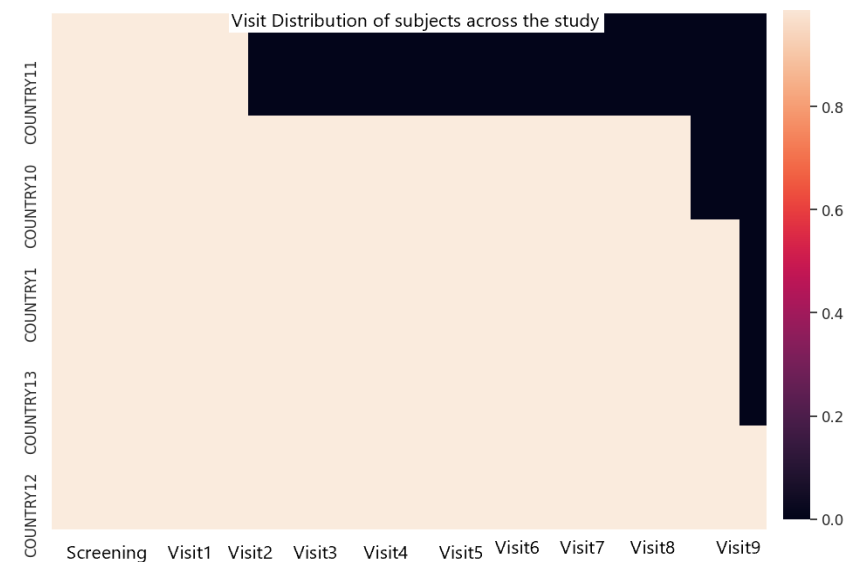
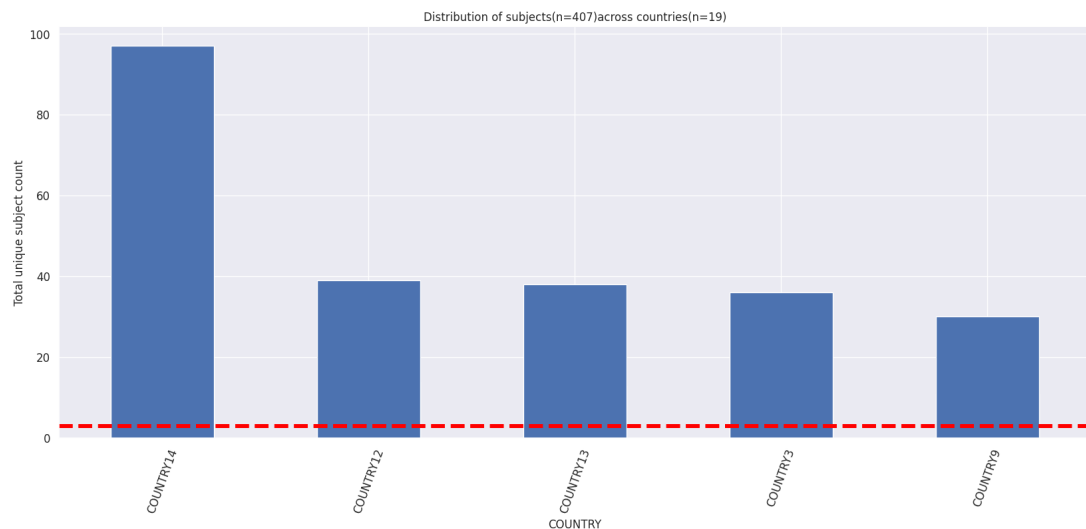


Country level use cases

- Initial data quality examination
- Countries impacted by Covid crisis
- Countries with AE/Disposition anomalies

Comparing study progression

By carefully examining initial data quality



Reviewing critical to:

- ✓ Subject enrollment counts
- ✓ Visit structure
- ✓ Site counts
- ✓ Measurement counts

Country Review functionality – study case 1

Country impacted by regional crisis

Total # Country High Risk

1

3.2%
% of Sites

Total # Country Medium Risk

3

9.7%
% of Sites

Total # Country Low Risk

27

87.1%
% of Sites

Country 1 was
signalized with
High risk



Country Identifier	Country Rank	Country Risk	Number of Sites	Number of Subjects	Percentage of High Risk Sites
Country 1	1	High	4	63	0.25
Country 2	2	Medium	7	85	0
Country 3	3	Medium	3	36	0
Country 4	4	Medium	4	46	0
Country 5	5	Low	2	13	0
Country 6	6	Low	16	104	0
Country 7	7	Low	5	21	0
Country 8	8	Low	1	7	0

Colors:

- Medium
- Low
- High

Country Identifier	Country Rank	Country Test Risk	Site Identifier	Signal Magnitude	Test Category	Test Identifier
COUNTRY1	1	High	SITE3	0	Adverse Events	TEST1
COUNTRY1	1	High	SITE4	0	Dispositions	TEST2
COUNTRY1	1	High	SITE5	0	Lab data	TEST3
COUNTRY1	1	High	SITE6	0	Lab data	TEST4
COUNTRY1	1	High	SITE1	0	Dispositions	TEST5
COUNTRY1	1	High	SITE2	0	Lab data	TEST6
COUNTRY1	1	High	SITE3	0	Adverse Events	TEST7
COUNTRY1	1	High	SITE4	0.01	Adverse Events	TEST8
COUNTRY1	1	High	SITE5	0.01	EDC data	TEST9
COUNTRY1	1	High	SITE6	0.01	Lab data	TEST10
COUNTRY1	1	High	SITE13	0.01	Lab data	TEST11
COUNTRY1	1	High	SITE14	0.01	Lab data	TEST12
COUNTRY1	1	High	SITE15	0.01	Questionnaire data	TEST13
COUNTRY1	1	High	SITE16	0.01	Adverse Events	TEST14

Country Identifier	Country Rank	Country Risk	Site Identifiers	Signal Summary	Test Identifiers	Percentage of Signal Occurring Sites
COUNTRY1	1	High	SITE15	test results for subject visits across time are more different from each other than expected	TEST1xPLASMA, TEST2xPLASMA, TEST3xPLASMA	100
COUNTRY1	1	High	SITE4	test results collected around the same calendar day at the site are more different from each other than expected	TEST7xLBLOOD, TEST8xLBLOOD	100
COUNTRY1	1	High	SITE3	test results for subject visits across time are more different from each other than expected	TEST9xBLOOD	100
COUNTRY1	1	High	SITE5, SITE6, SITE13, SITE14	test results collected around the same calendar day at the site are more alike than expected	TEST3, TEST4, TEST11, TEST12	75
COUNTRY4	9	Low	SITE38	test results collected around the same calendar day at the site are more different from each other than expected	TEST23, TEST27	100

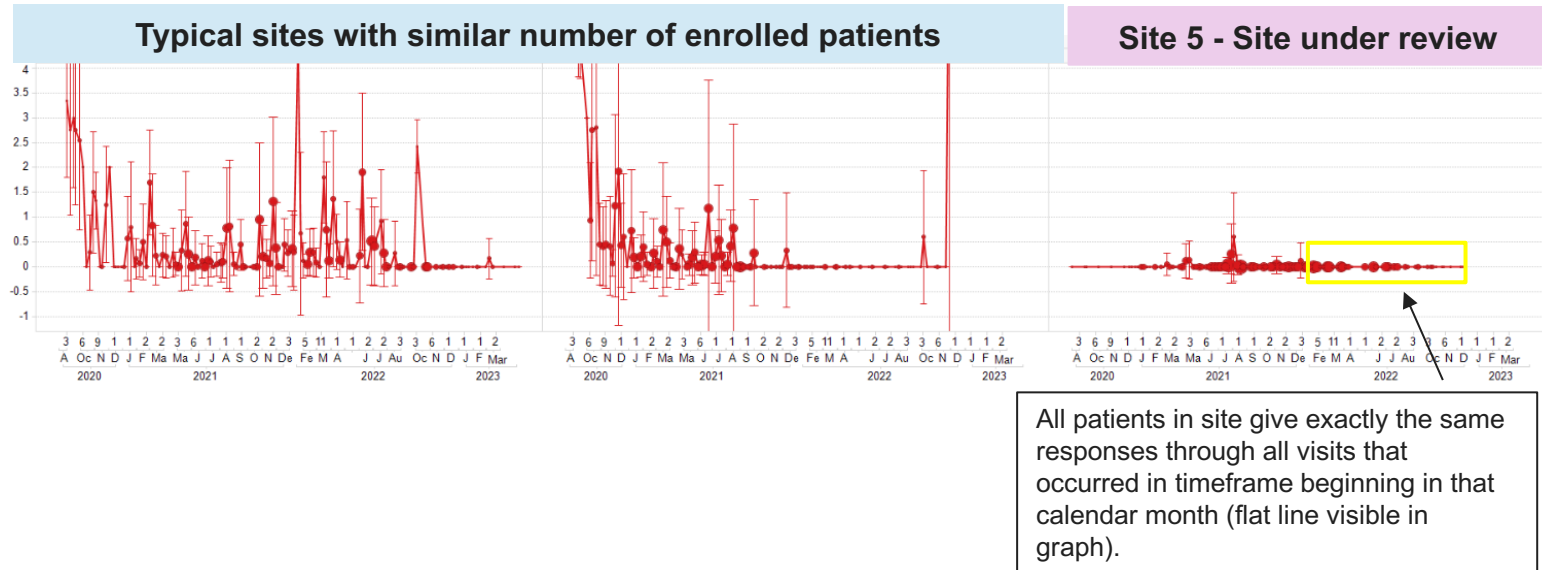
Mark items to view details here.

Sites from
High-risk
Country 1
grouped by the
same signal
occurrences

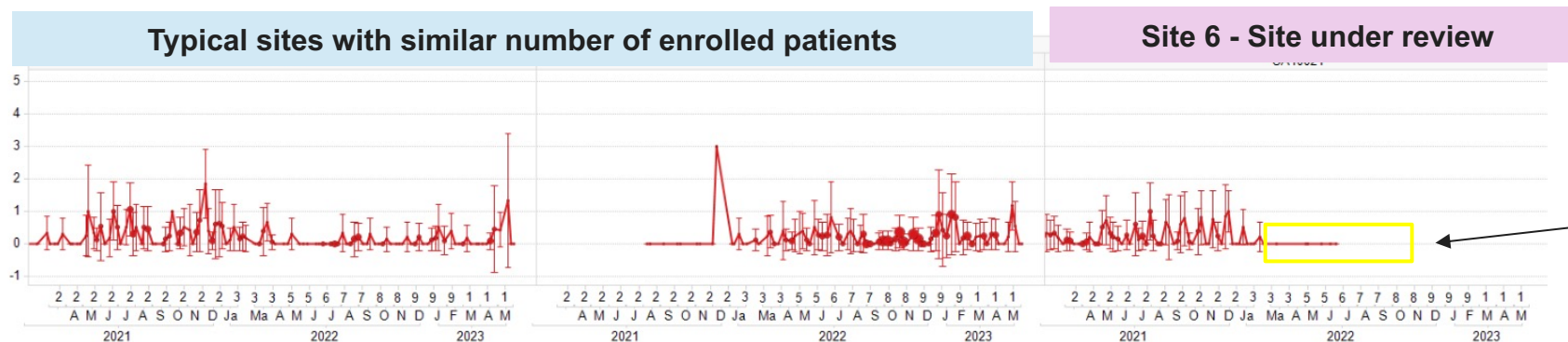


Country Review functionality – study case 1

Timeline of data quality impact



Site 6
TEST3
test identifier

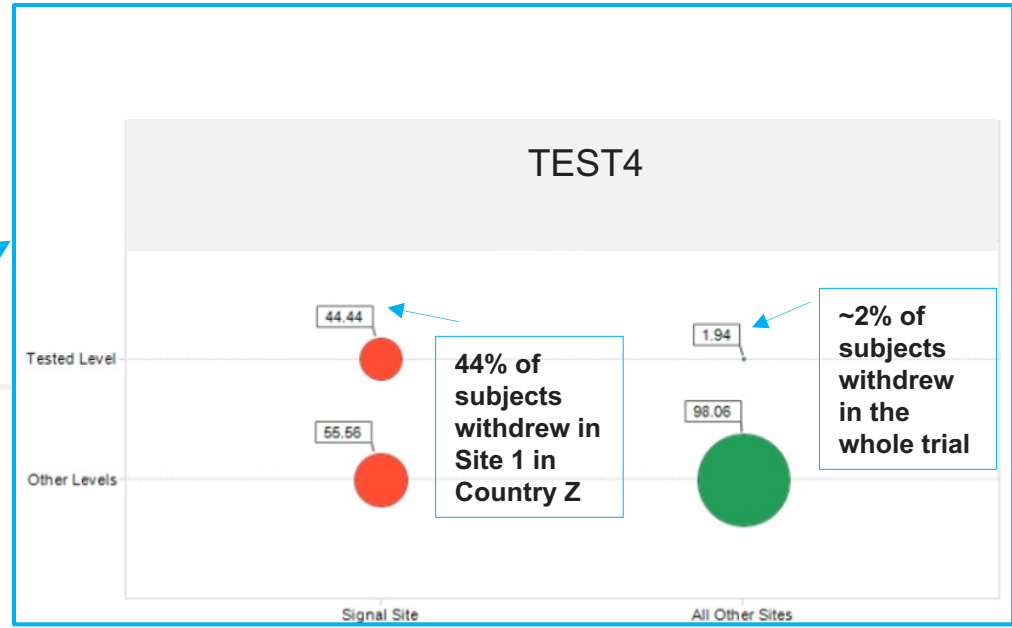


Categorical Data Dispositions—case in Country Z



Site Identifier	Site Rank	Site Risk	Number of Subjects
Site 1 in Country Z	2	High	9
Site 1 in Country X	9	Medium	10
Site 1 in Country Y	44	Low	5
Site 1 in Country F	76	Low	4
Site 1 in Country G	83	Low	6

Site Signal Panel				
Site Rank ▲	Site Risk	Site Identifier	Signal Summary	Test Identifiers
2	High	0007	Statistical Anomaly (occurrence in more subjects than expected) in TEST4	subjects with TEST4



Site 1 in Country Z

Footnote: **CMWG* Meeting:** The abbreviation "CMWG" stands for Central Monitoring Working Group. Central monitoring managers extensively review clinical and operational data signals/trends/outliers that indicates potential risks and present it to the central study teams in CMWG meetings; necessary actions resulting from these discussions are subsequently followed up for corrective/preventive actions.

Categorical Data Follow up– study case in Country Z

SCENARIO

- CSS run flagged a signal* for a statistical anomaly for 1 site in **Country Z**.
- The anomaly flagged a site with 44% discontinued with reason withdrawal by subject (trial level rate 2%)
- Central monitoring managers deep dive revealed a **country specific signal**

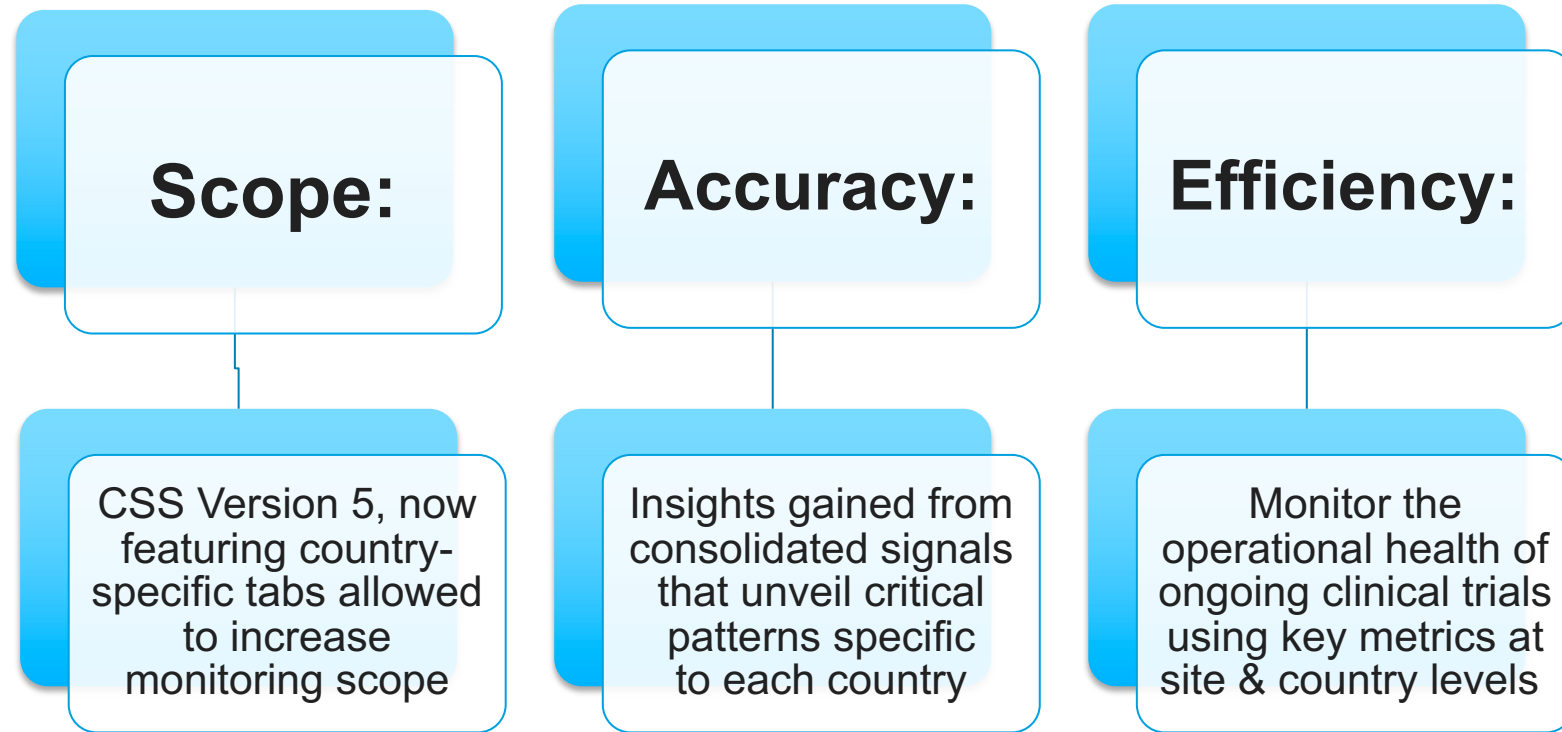
CAUSE

Main contributing factors:

- **Processes related:**
Study/protocol expectations may not have been clearly set with subjects
- **Lack of clarity** or issues with consent withdrawal process.

ACTIONS

- Sites encouraged to gather and **update eCRF**
- **Site Communication letter** shared with all sites stressing the importance of data quality.
- **Patient retention training** being revisited and delivered for all key risk sites.



Future ideas:
Further Automation: Real time study data refreshes for monitoring insights

Across portfolio: Comparison between studies to review portfolio level insights

Predictive modelling: Risk analysis proactively highlighted using data driven trends

Acknowledgments

Many thanks to Md Umarfaruque, ZS, Janssen R & D, Weronika Stepczyk, Janssen R & D, Warsaw, Poland, Susan Meier, Janssen R & D, PA, USA, Marion Wolfs, Janssen R & D, NL, Kris Lauwers, Janssen R & D, Beerse, Belgium, Bhargava Reddy, Janssen R & D, PA, USA for their contributions and timely suggestions.

*Many thanks to all Janssen R&D, J&J; Especially all team members of Advanced insights, Data science, Central Monitoring team, Data Management teams
Special thanks to Central Data science platform team (Domino) and
R&D Advance Computing- Data Analytics & Engineering teams
FLEX PRAHS/ICON PLC/Zeus*

Many thanks PHUSE EU CONNECT 2023 for this opportunity



1) **Lauwers et al, 2018**: Central Statistical Surveillance in Clinical Trials. Kris Lauwers, Stefan Van Yper, Giordano Lo Regio, Janssen R & D

https://phuse.s3.eu-central-1.amazonaws.com/Archive/2018/Connect/EU/Frankfurt/PAP_AR04.pdf

2) **Nayak et al, 2021**: Programming infrastructure to automate process flow for Central Statistical Surveillance. Archana Nayak, Salvatore Saporito, Kris Lauwers, Stefan Van Yper, Mateusz Hinzmann, Bhargava Reddy, Janssen R & D

https://phuse.s3.eu-central-1.amazonaws.com/Archive/2021/Connect/EU/Virtual/PRE_AR03.pdf

3) **George et al, 2023**: Quantifying the impact of the COVID-19 pandemic on cancer center clinical trial operations <https://pubmed.ncbi.nlm.nih.gov/37467065/>