



IC-10

Collaborative Development of a Clinical Trial Data
Anomaly Detection Tool within an Industry
Consortium

PHUSE/FDA Innovation Challenge

Pekka Tiikkainen
February 2024

The Inter coMPany quALity Analytics (IMPALA)

Agenda

- Central Statistical Monitoring
- IMPALA
- Clinical Trial Anomaly Spotter (CTAS)
- Examples of anomalies
- Our vision for CTAS

Agenda

- Central Statistical Monitoring
- IMPALA
- Clinical Trial Anomaly Spotter (CTAS)
- Examples of anomalies
- Our vision for CTAS

Central Statistical Monitoring (CSM)

- Trial sponsor is responsible for ensuring the high quality of the data collected
- Examples of sponsor activities
 - On-site visits: e.g. source data verification, review of documents, investigator interviews
 - **Central *statistical* monitoring:** analysis of results collected at the site to identify anomalies and outliers. Are vital signs measured at the site *systematically* larger than in other sites? Are the values for a local lab replicated? Have the visits been taken at correct intervals? Were inclusion and exclusion criteria met?
- Data is reviewed regularly as the study is on-going.
 - Certain errors can be corrected by site, e.g. data entry errors.
 - If a systematic bias at a site is identified early enough, the site can be offered additional training to ensure compliance.
 - Fraud performed by site can lead it to be excluded from the study.



Motivation for CSM

- 100% SDV (source data verification) is a very inefficient process as less than 1.7% of data gets changed due to queries [1]. Also, a significant part of the study budget is spent on SDV [2].
- With CSM, only signals would lead to an action, saving resources for both the sponsor and the investigator.
- CSM is also encouraged by the FDA [3].
- In addition, CSM can highlight more complex data patterns that would have been missed otherwise – leading to improved data quality [4].



[1] Stokman, P. G. et al. (2021) “Risk-based Quality Management in CDM An inquiry into the value of generalized query-based data cleaning”, Journal of the Society for Clinical Data Management

[2] Funning S. et al. (2009) “Quality assurance within the scope of Good Clinical Practice (GCP)—what is the cost of GCP-related activities? A survey within the Swedish Association of the Pharmaceutical Industry (LIF)’s members”, The Quality Assurance Journal.

[3] Guidance for Industry - Oversight of Clinical Investigations — A Risk-Based Approach to Monitoring. Docket Number FDA-2011-D-0597.

[4] de Viron, S. et al. (2023) “Does Central Monitoring Lead to Higher Quality? An Analysis of Key Risk Indicator Outcomes”, Therapeutic Innovation & Regulatory Science.

Agenda

- Central Statistical Monitoring
- **IMPALA**
- Clinical Trial Anomaly Spotter (CTAS)
- Examples of anomalies
- Our vision for CTAS

IMPALA consortium

- Industry consortium IMPALA has been formed to develop tools, provide training and create best practices for quality in clinical trials.
- IMPALA aims to engage with Health Authorities inspectors on defining guiding principles for the use of advanced analytics to complement, enhance and accelerate current QA practices.
- Another important activity is co-development of open-source AI/ML tools for QA.
- For more information on the IMPALA consortium, please visit <https://impala-consortium.org/>



AstraZeneca

Galápagos

sanofi

Roche

astellas

Lilly

AMGEN

Bristol Myers Squibb

Pfizer

MSD

Boehringer
Ingelheim

NOVARTIS

Biogen

J&J

MERCK

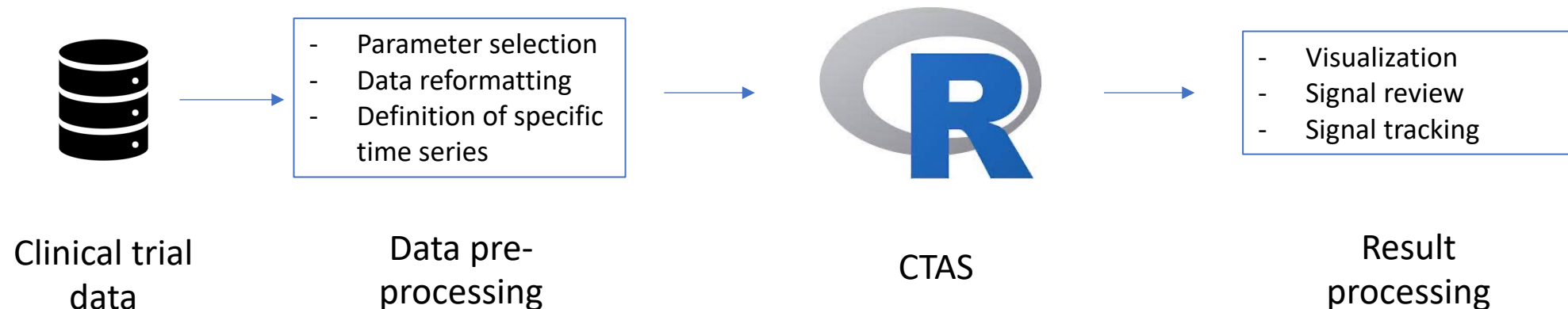
BAYER

Agenda

- Central Statistical Monitoring
- IMPALA
- **Clinical Trial Anomaly Spotter (CTAS)**
- Examples of anomalies
- Our vision for CTAS

Clinical Trial Anomaly Spotter (CTAS)

- CTAS (Clinical Trial Anomaly Spotter) is a CSM tool focused on identifying anomalous time series of continuous variables.
- “Unsupervised” tool which does not use pre-specified KRIs. The identified anomalies can be such that were not anticipated during study design.
- Originally developed by Bayer where it is available for all company studies.
 - The Bayer-internal tool was presented in PHUSE EU Connect 23. Please see refs. [1] and [2] for details.
- Code base shared with IMPALA for co-development in July 2023.



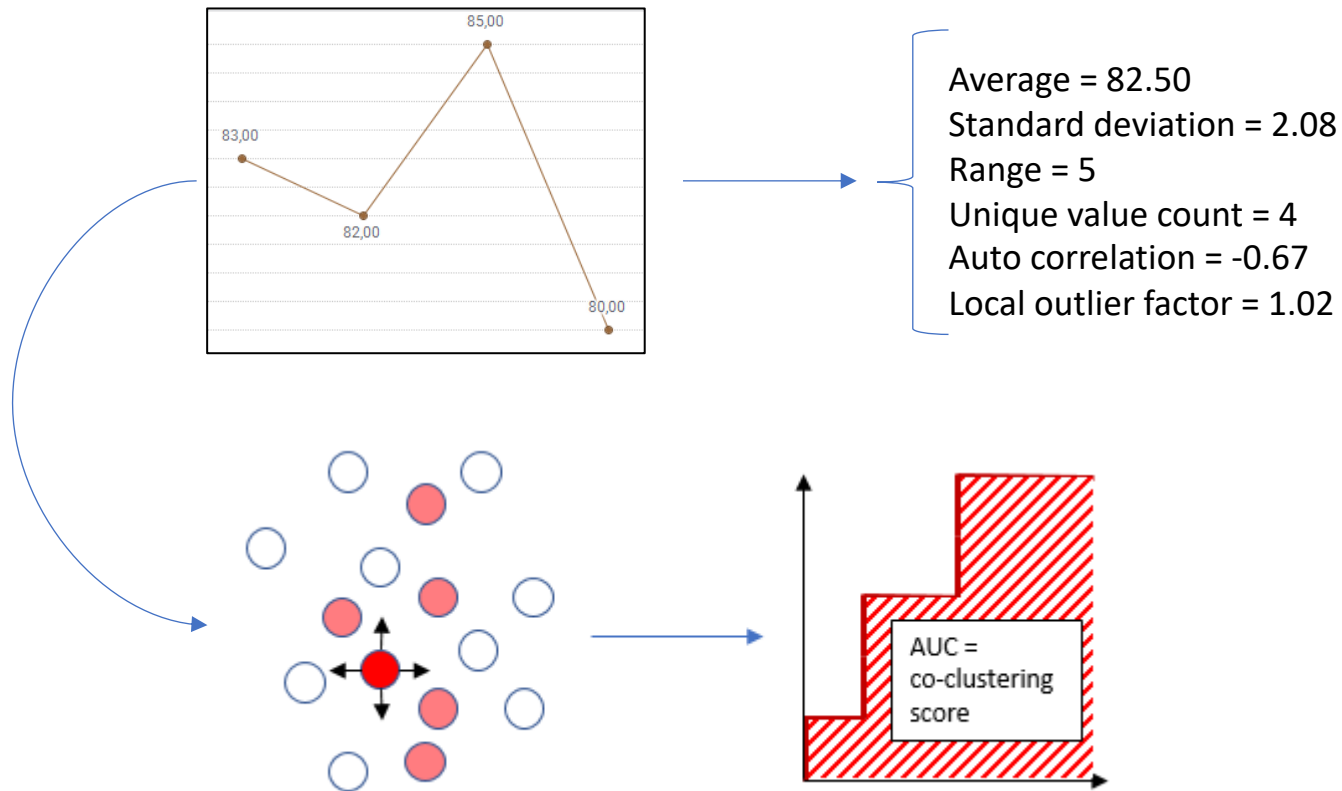
[1] https://phuse.s3.eu-central-1.amazonaws.com/Archive/2022/Connect/EU/Belfast/PRE_AR04.pdf

[2] https://phuse.s3.eu-central-1.amazonaws.com/Archive/2022/Connect/EU/Belfast/PAP_AR04.pdf

CTAS algorithm

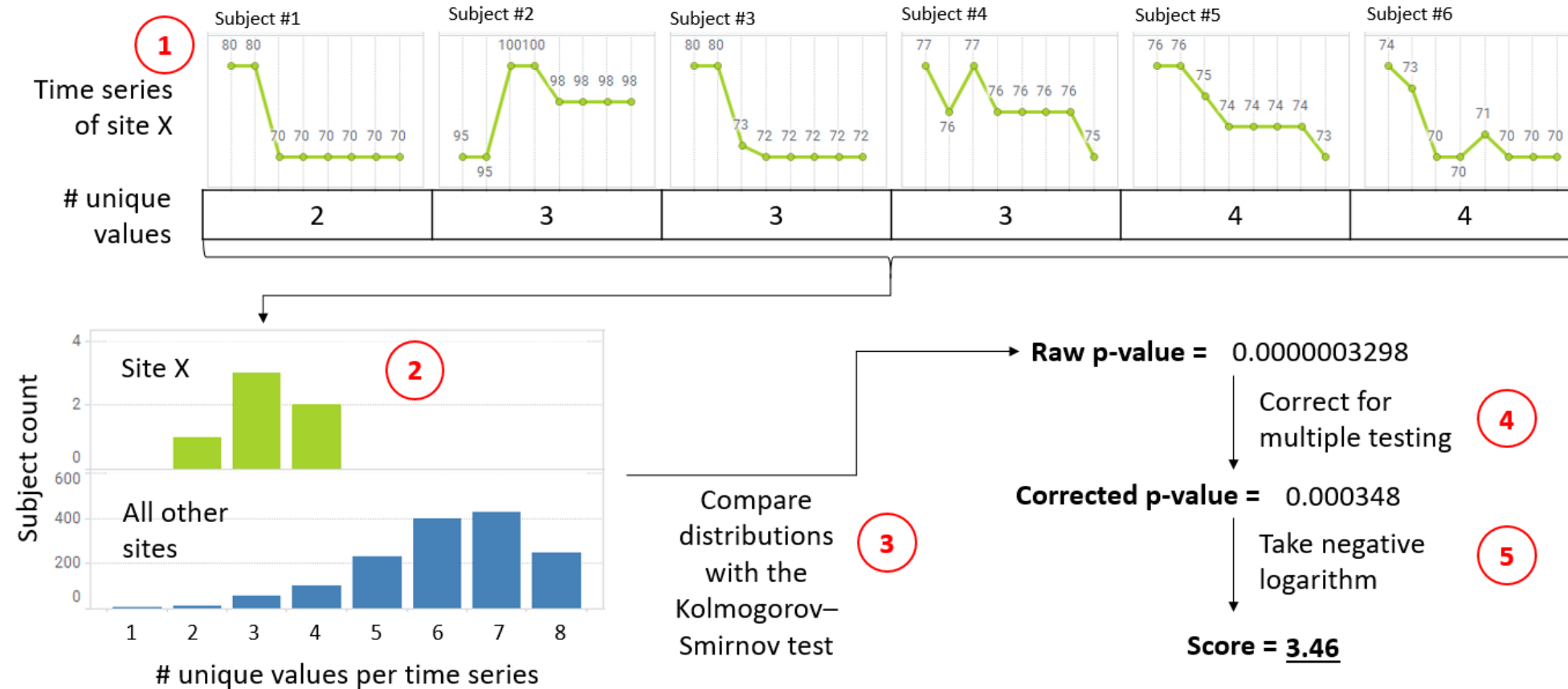
Time series features

Each time series is summarised as a set of features.



CTAS algorithm

Flagging sites



What CTAS does **not** cover (at least not yet)

- Categorical variables
- Time series forecasting
- Detection of **individual** anomalous data points in time series
- Key risk identifiers (KRIs)
 - There are other tools which cover these, e.g. Good Statistical Monitoring (gsm) package (<https://gilead-biostats.github.io/gsm/>)
- Data pre- and post-processing
- Visualization of results
- CTAS is an AI tool utilizing statistical testing, with a potential modeling approach under evaluation; its adoption hinges on performance enhancement.



CTAS is open source

After few months of co-development, IMPALA published CTAS as an open-source R package in December 2023.

The source code is available on the public IMPALA Github repository under the MIT License:
<https://github.com/IMPALA-Consortium/ctas>

Everyone is invited to test and use the package and report bugs and/or suggest new features.

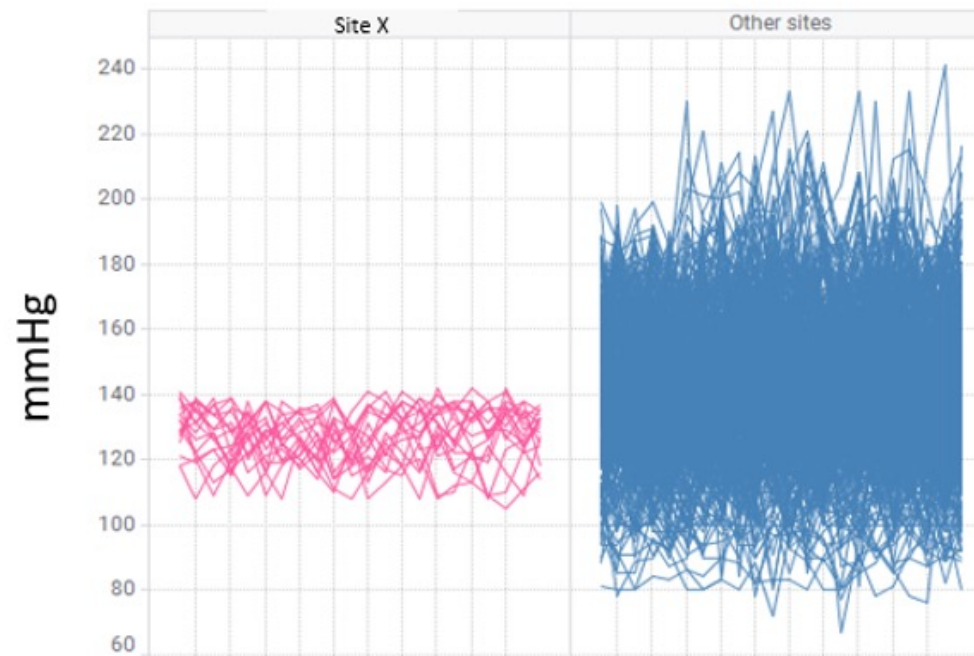
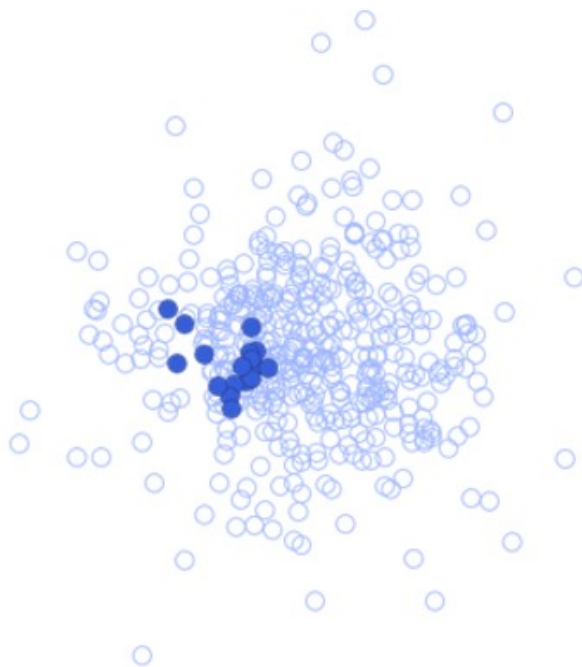


Agenda

- Central Statistical Monitoring
- IMPALA
- Clinical Trial Anomaly Spotter (CTAS)
- **Examples of anomalies**
- Our vision for CTAS

Examples of anomalies identified with CTAS

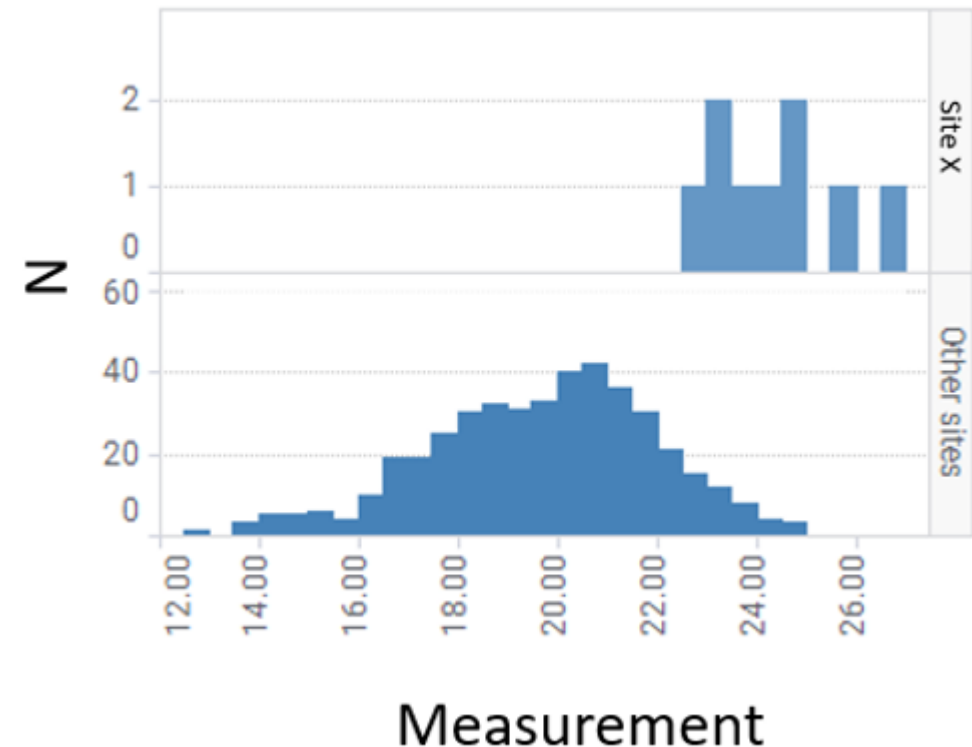
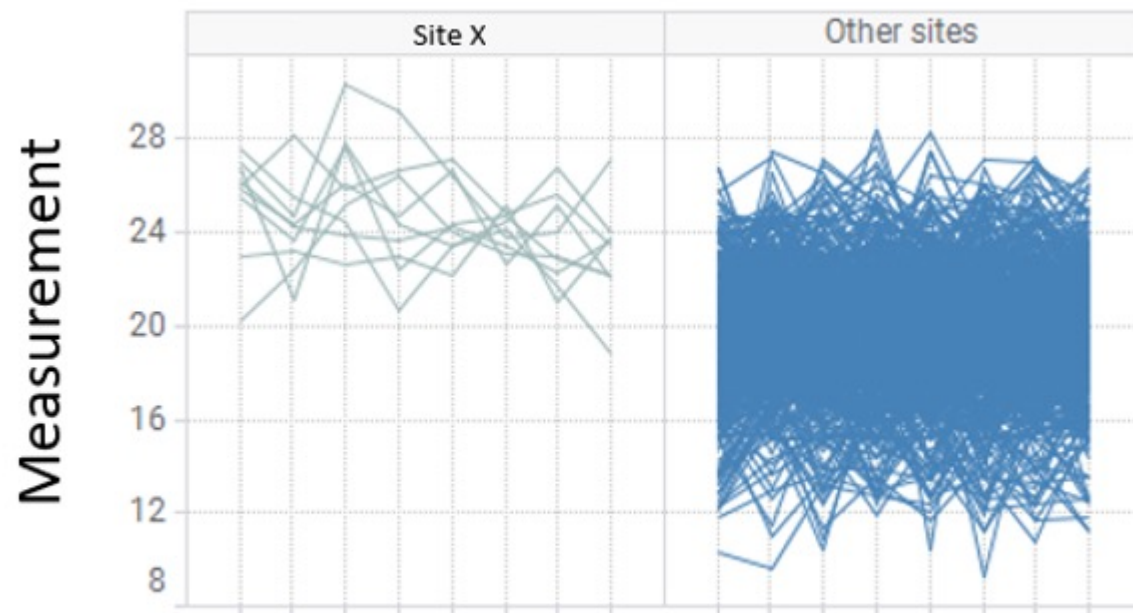
Site with co-clustered blood pressure profiles (Bayer)



Potential reasons: result fabrication, “sample” splitting, bias in subject selection

Examples of anomalies identified with CTAS

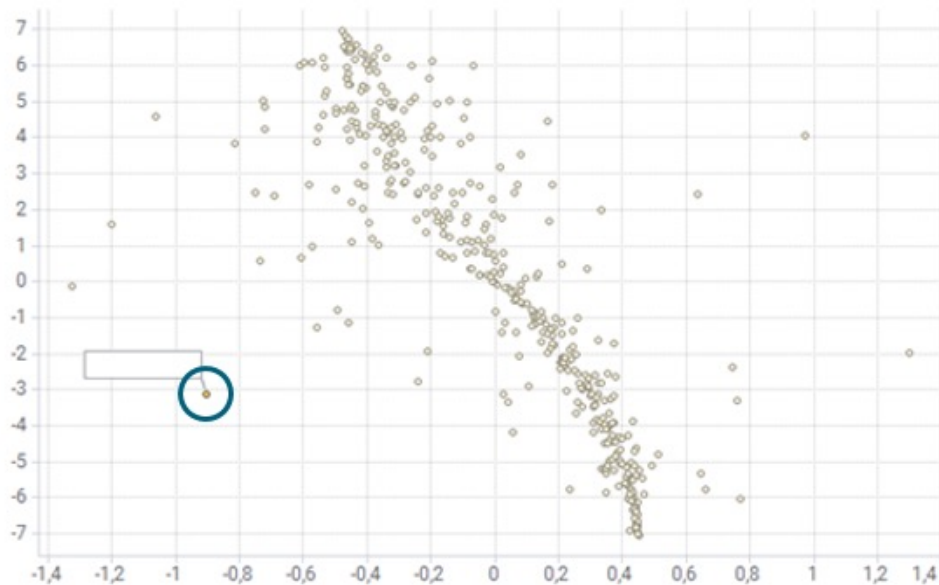
Site with unusually large laboratory assay results (Bayer)



Potential reasons: bias in subject selection, in-correct assay calibration (local labs), errors in sample handling (central labs)

Examples of anomalies identified with CTAS

Subject-level anomaly: “interesting” weight profile (Bayer)



Reason: data entry error

Agenda

- Central Statistical Monitoring
- IMPALA
- Clinical Trial Anomaly Spotter (CTAS)
- Examples of anomalies
- Our vision for CTAS

Our vision for CTAS

- Our vision: CTAS to become an industry best practice for central statistical monitoring adopted by both the industry and health authorities.
- Authorities, like FDA, are invited to test the tool and actively contribute to its further development.
- The best way to influence the tool development would be to join IMPALA as a health authority member.

Roche testimonial:

Roche is exploring the use of the Clinical Trial Anomaly Spotter (CTAS) to have better oversight of clinical data integrity from a quality assurance perspective and thereby support our audit site selection and audit scope. Initial results show that we can successfully detect questionable data entries in critical variables. We believe that the use of open source quality analytics has the potential to transform the way audits are conducted and will raise the bar for the quality of clinical data.

Boehringer Ingelheim testimonial:

Boehringer Ingelheim is actively investigating Clinical Trial Anomaly Spotter (CTAS). We believe that detecting anomalous data patterns can augment the approach of data assisted risk assessment in clinical trials. This approach could lead to quicker decision-making, better data quality, and faster development of breakthrough therapies to transform patient's lives, the essence of Boehringer Ingelheim's innovative spirit."



Thank you!

Author contact information:

Pekka Tiikkainen
Bayer Pharmaceuticals
pekka.tiikkainen@bayer.com

