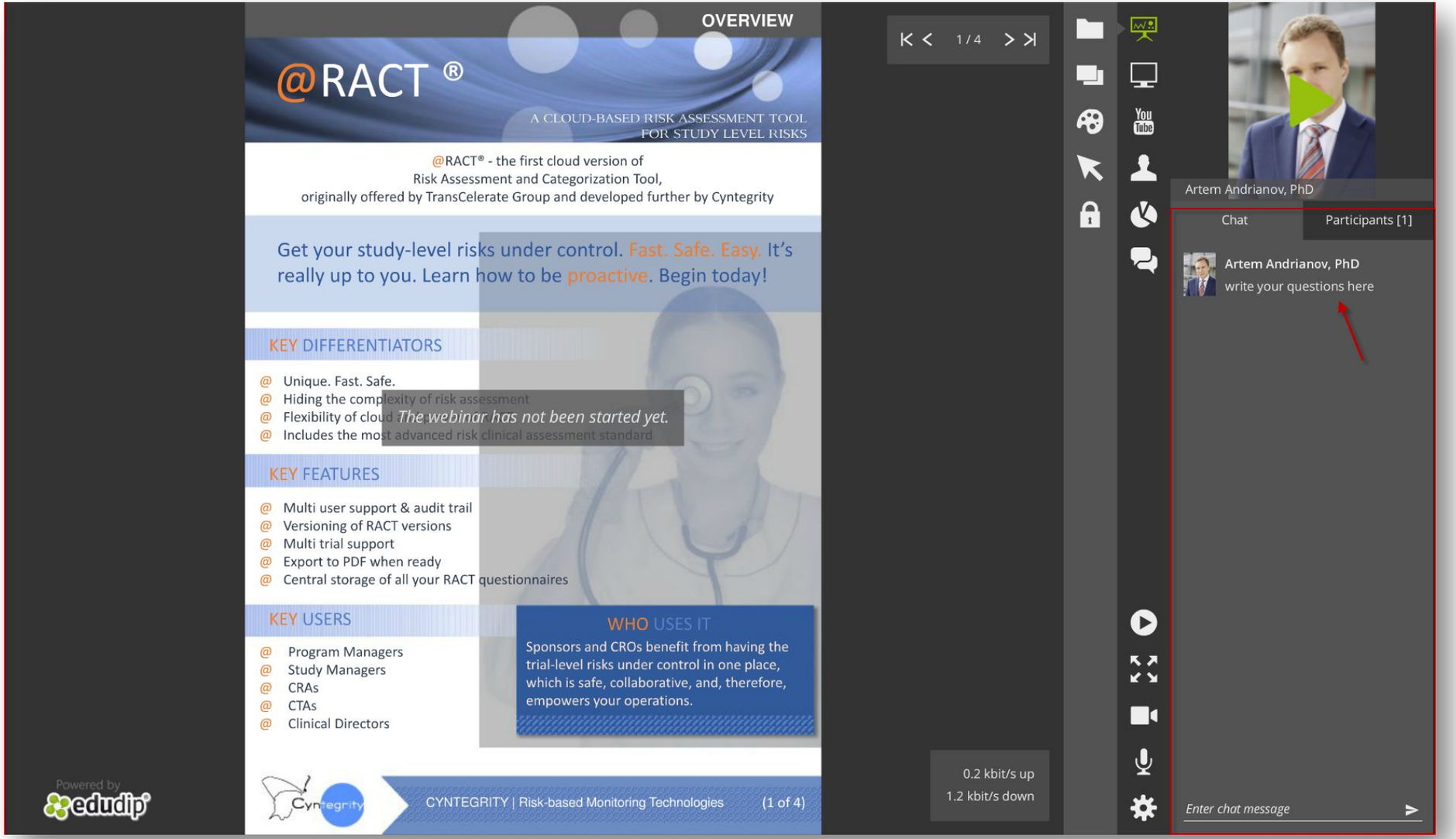
Trust in Data

Fraud and Sloppiness Detection in Clinical Trials

With Practical Examples

Webinar | Thursday, September 26, 2019

Ask your questions here >



The Company

Since:	**2014**
Founders:	RB(Q)M pioneers, **ERT**'s former Quality Data Systems Development Lead and team and co-founded by **Bayer**'s former Head of Global Data Management
Focus:	Full-service RB(Q)M solutions, exclusively **designed for clinical purpose**
Headquarters:	Olof-Palme-Str. 15, D-60439 Frankfurt am Main, **GERMANY**

Artem Andrianov, PhD

CHIEF EXECUTIVE OFFICER

Artem Andrianov has more than 15 years of experience in data quality projects (GlaxoSmithKline, Johnson & Johnson and ERT), RB(Q)M and data management for the Life Sciences and Medical Devices industries.

Artem is a much sought-after, charismatic speaker on topics related to RB(Q)M, Spirometry, ECG, Oncology, ePROs and Clinical Data Management at conferences such as DIA, PCT, DGGF and PharmaDay.

Artem holds a PhD in Mathematics and master's degrees in Business Administration, Financial Management and Corporate Strategy.

The Butterfly Effect: Why Small Acts Can Have Big Effects

Fraud and Sloppiness Detection in Clinical Trials

Definitions US Public Health Service:

- *"Research misconduct means **fabrication**, **falsification**, or **plagiarism** in proposing, performing, or reviewing research, or in reporting research results;*
- *Fabrication is **making up** data or results and recording or reporting them;*
- *Falsification is **manipulating** research materials, equipment, or processes, or changing or omitting data or results such that the research is not accurately represented in the research record;*
- *Plagiarism is the **appropriation** of another person's ideas, processes, results, or words without giving appropriate credit;*
- *Research misconduct **does not include honest error or differences of opinion**."*

Do you want to be the next 'scandal'?

- *More than **40% of researchers** are aware of misconduct but do not report it.*[1]
- *More than **2000 scientific articles** have been retracted over the last 40 years.*[2]
- *The number of retractions has **increased dramatically in recent years** and most of these retractions are due to **research misconduct**, especially **data fraud**.*[3]

What's your organization doing to protect against fraudulent activity? Have you taken the necessary measures to reduce **monetary loss**, keep **brand reputation** high while keeping organizational efficiencies on track?

1. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3700330/
2. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4340084/
3. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4340084/

Either intentionally or unintentionally, fraud and error in clinical research happen…

Why Fraud or Misconduct Happens?

Why Fraud and Misconduct Happens?

The reasons:

- Need to obtain a desired result, e.g. ‘statistical significance’
- Monetary gain, enhancement of prestige
- Compensate for laziness, sloppiness in data collection
- Include subjects who would otherwise be excluded
- Ignorance, naivety
- Carelessness, sloppiness
- Improper methods
- Statistical ‘fallacies’

Source: Habermann B, Broome M, Pryor ER, Ziner KW. Research coordinators’ experiences with scientific misconduct and research integrity. Nurs Res. 2010;59:51–7.

Why Fraud and Misconduct Happens?

The consequences:

- Damages corporate reputation - fraud cases can destroy a large research organization
- Financial costs connected with recalls
- Damage reputation of many innocent researchers and collaborators
- Negative impact to related on-going research

➔ Victims

Why Fraud and Misconduct Happens?

The different types:

SITES

- Manipulation with protocol
- Misuse of measurement tools
- Wrong calibration
- Inclusion criteria fraud
- Manipulation of parameters for trial inclusion

MEASUREMENTS

'**Professional patients**' – people who participate in many trials simultaneously.

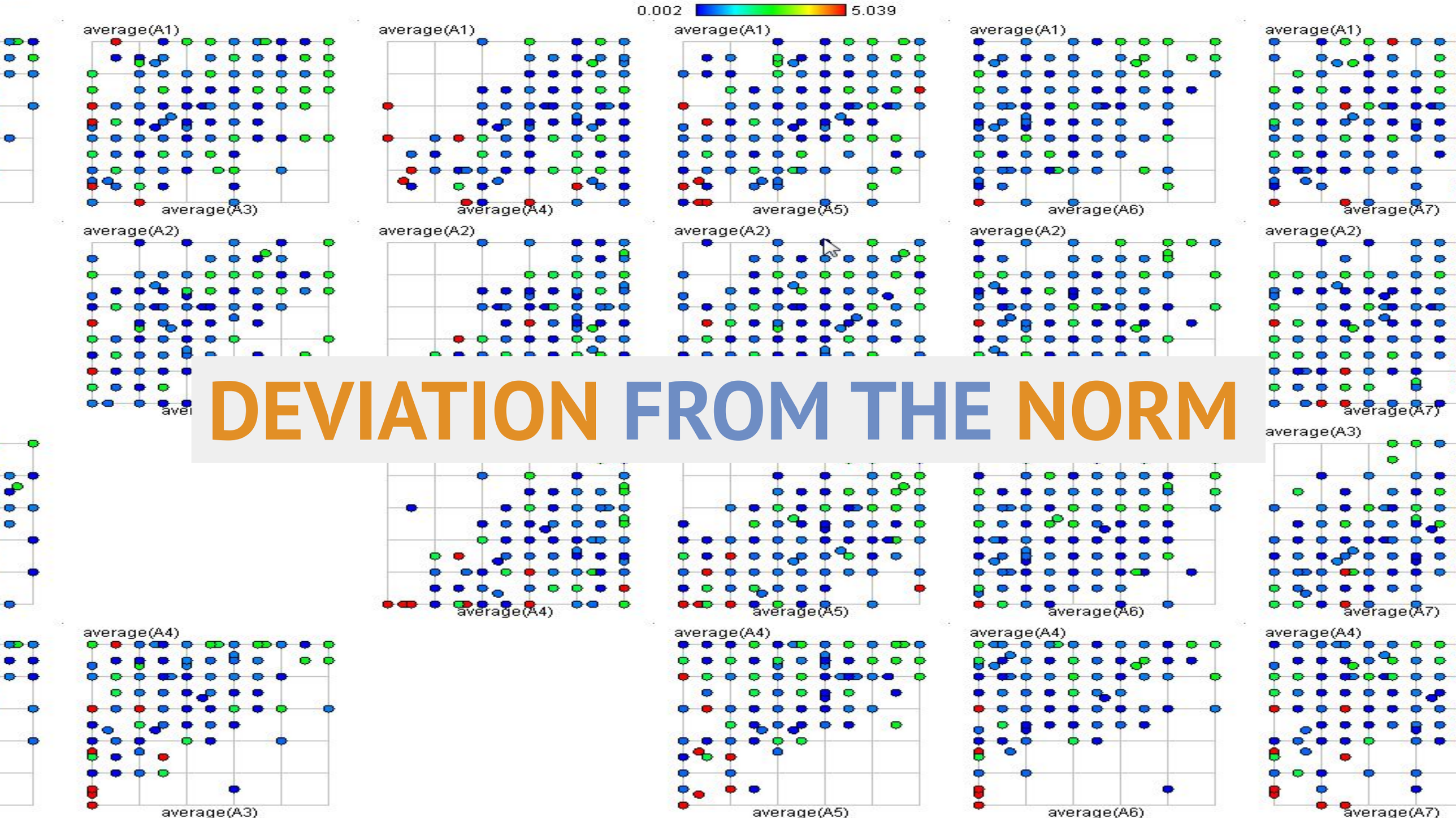
CROs

- Error-hiding
- Unintended technical defects
- Unnoticed data defects

DATA EXPORTS

- Error-hiding
- Inventing new data (**forging**)
- Reporting only selected observations (**cooking**)
- Deleting outliers (**trimming**)

How to Detect Fraud and Error?


0.002
5.039
average(A1)
average(A3)
average(A1)
average(A4)
average(A1)
average(A5)
average(A1)
average(A6)
average(A1)
average(A7)
average(A2)
average(A2)
average(A2)
average(A2)
average(A2)
average(A7)
DEVIATION FROM THE NORM
average(A3)
average(A4)
average(A5)
average(A6)
average(A7)
average(A4)
average(A4)
average(A4)
average(A4)
average(A3)
average(A5)
average(A6)
average(A7)

MyRBQM® Analytics and Fraud Detection

MyRBQM® Analytics and Fraud Detection

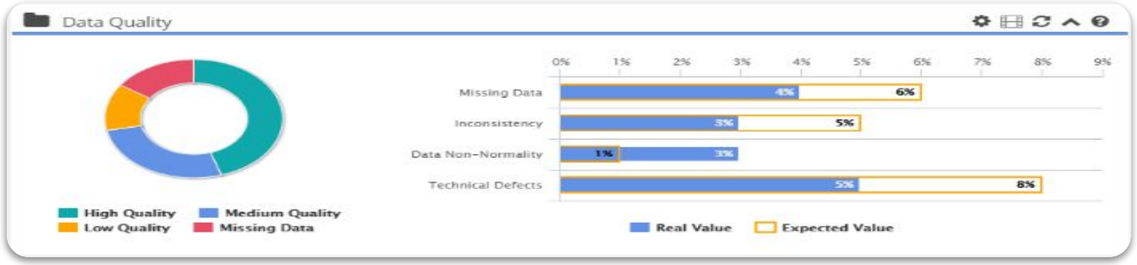
Technology essential features:

Clinical Risk Management

- Umbrella Principle – unification of CTMS, EDC, etc.
- Risk mitigation communication
- Concentration on systemic risk, not random artefacts

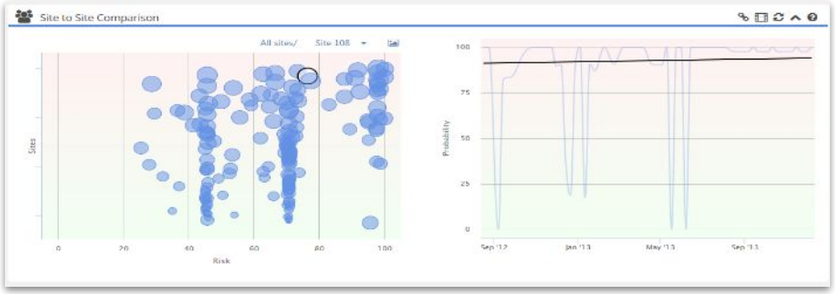
Data Quality Monitoring

- P-value control for essential variables
- Geographical quality spread
- Data quality dynamics and comparison values



Site Ranking

- Transparent communication and site involvement
- Historical Root-Cause feedback graph



MyRBQM® Analytics and Fraud Detection

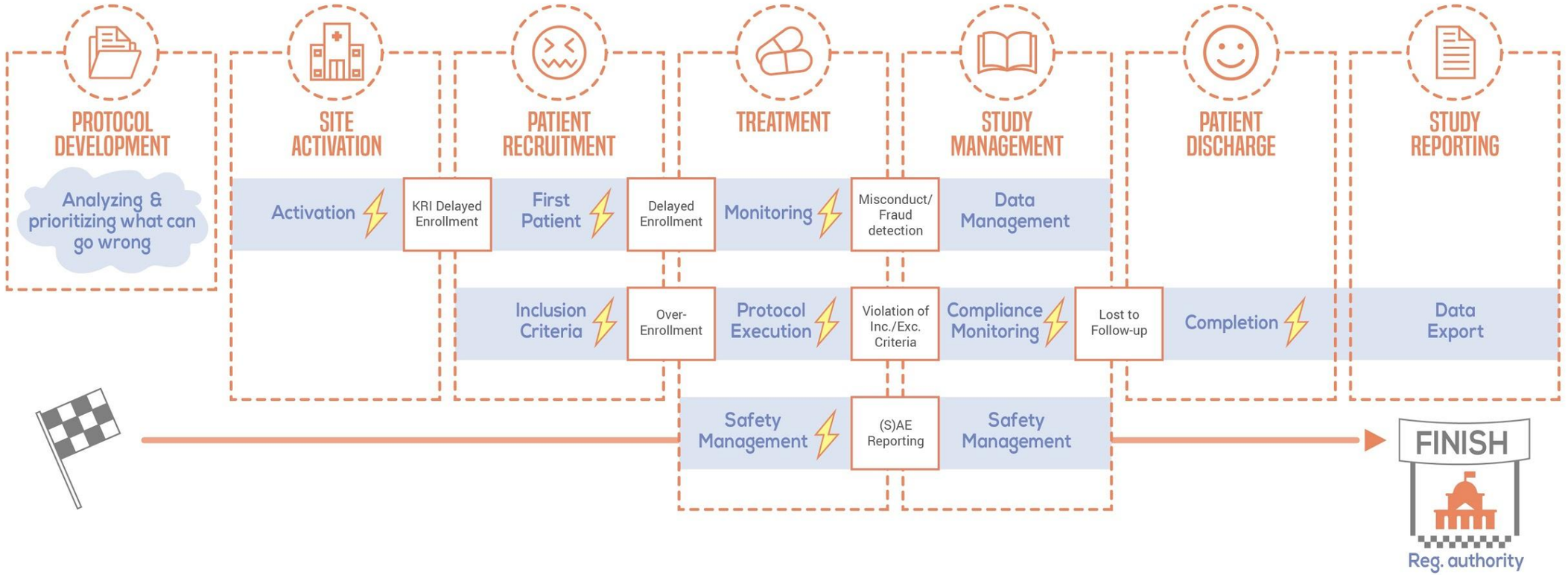
Preconditions:

Regulatory compliance - tools and technology designed for clinical purpose

MyRBQM® Analytics and Fraud Detection

Preconditions:

Holistic approach - all risks in scope from start to finish

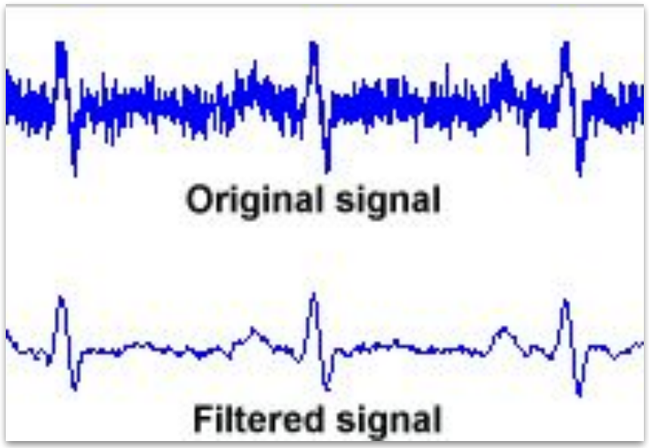


MyRBQM® Analytics and Fraud Detection

Key Risk Indicators (KRIs):

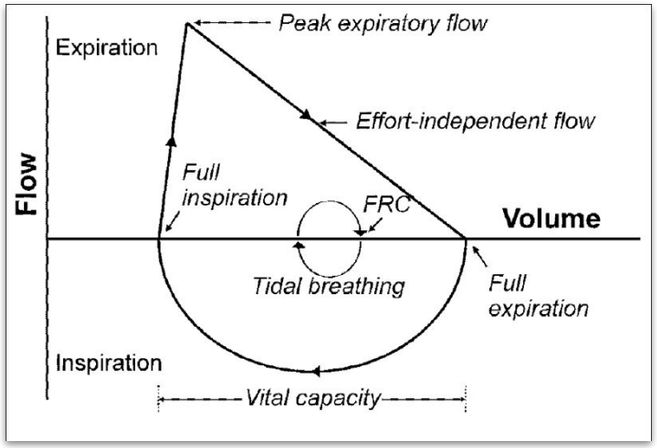
ECG KRIs

- ECG Noise Level
- Lead Misplacement
- Empty leads
- Low Contact Impedance
- Parameter Analysis for, e.g., QT check



Spirometry KRIs

- ATS Criteria Check
  - Repeatability
  - Recognizing correct and incorrect FVC maneuvers
- Efforts quality
- Parameter Integrity



How Can We Prevent Fraud and Misconduct?

How Can We Prevent Fraud and Misconduct?

The measures:

- Adopt **zero tolerance** - all suspected misconduct
- Using **statistical and data mining tools** in identification of inconsistent data
- Using **centralized statistical monitoring methods** (across study and sponsor)
- Continuous acceptability & sanity **data checks**
- **Rules Engine Check** (based on medical knowledge)
- **Data mining**: Gaining new information from data analysis and using it to tune the detection
- Development of **fraud classification** and a method of fraud **categorization**
- Calculating and assigning independent **Fraud Credibility Ranking** of Sites, CROs and organizations involved in clinical research (Low, Medium, High Risk Site)
- More discussion should be directed towards possibilities to decrease **incentives** connected with fraud, misconduct and sloppiness

MyRBQM® CSM Fraud Detection Algorithms

MyRBQM® Centralized Statistical Monitoring (CSM)

Fraud detection algorithms:

Digit Preference - within-study comparison

The algorithm is able to **compare leading or trailing digits** between user-defined groups in a clinical trial. For example, for sites or for randomization arms (for checking whether randomization is correct).

The algorithm:

- calculates a contingency table for frequency distribution of digit of interest
- performs Cochran-Mantel-Haenszel “Row Mean Square Differ” test, which effectively compares rows of contingency tables between each other
- Scores are calculated from the p-values and are highlighted in a report if a p-value is lower than a user-defined significance level, **which means that the data is suspicious and the case should be investigated**

MyRBQM® Centralized Statistical Monitoring (CSM)

Fraud detection algorithms:

Digit Preference - What is Benford's Law?

Benford's law, also called the Newcomb–Benford law, the law of anomalous numbers, or **the first-digit law**, is an observation about the frequency distribution of leading digits in many real-life sets of numerical data. The law states that in many naturally occurring collections of numbers, **the leading significant digit is likely to be small**.

For example, in sets that obey the law, the number 1 appears as the leading significant digit about 30% of the time, while 9 appears as the leading significant digit less than 5% of the time. If the digits were distributed uniformly, they would each occur about 11.1% of the time. Benford's law also makes predictions about the distribution of second digits, third digits, digit combinations, and so on.

https://en.wikipedia.org/wiki/Benford%27s_law

MyRBQM® Centralized Statistical Monitoring (CSM)

Fraud detection algorithms:

Digit Preference - Benford’s Law

The algorithm is able to detect fraud by **checking if the first significant digit (FSD) frequency distribution complies with Benford's Law**. The algorithm triggers a warning if some of the data compliance properties are not observed.

This algorithm:

- finds frequencies of values of one or two leading digits in data
- finds second-order frequencies of two leading digits
- compares these frequencies with the expectation of Benford’s Law
- Scores are calculated from p-values and are highlighted in a report if a p-value is lower than a user-defined significance level, which means that the **data is suspicious and the case should be investigated**

MyRBQM® Centralized Statistical Monitoring (CSM)

Fraud detection algorithms:

Digit Preference - uniformity check

The algorithm detects data collection errors and fraud by **checking if the trailing digits frequency distribution is uniform or not**.

This algorithm:

- finds frequencies of values of one or two trailing digits in data
- compares these frequencies with the expectation of uniform distribution
- p-values produced by statistical tests are adjusted for multiple comparisons by the Benjamini-Yekutieli method and compared with the significance level specified by a user
- Scores are calculated from p-values and are highlighted in a report if a p-value is lower than a user-defined significance level, which means that the **data is suspicious and the case should be investigated**

MyRBQM® Centralized Statistical Monitoring (CSM)

Fraud detection algorithms:

Interquartile Range - IQR

IQR is used in case of univariate **outliers detection** to identify the values of numerical variables which are **abnormally distant** from most of the observations to detect **data collection errors, sloppy data or fraud.**

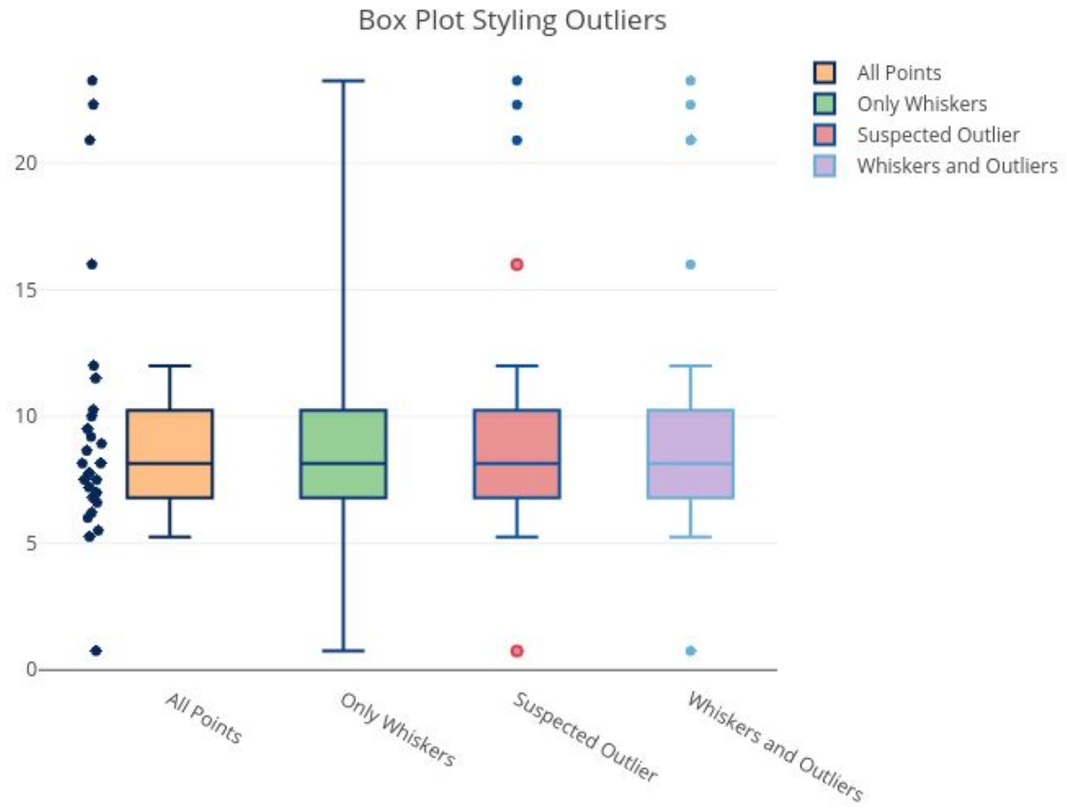


Image Source: https://plot.ly/r/box-plots/

MyRBQM® Centralized Statistical Monitoring (CSM)

Fraud detection algorithms:

Scatter Plot

A scatter plot is a two-dimensional data visualization that uses dots to represent the values obtained for two different variables — one plotted along the x-axis and the other plotted along the y-axis. The scatter diagram graphs pairs of numerical data, with one variable on each axis, to look for a relationship between them.

It helps to **detect fraud by observing correlations between variables and finding outliers**.

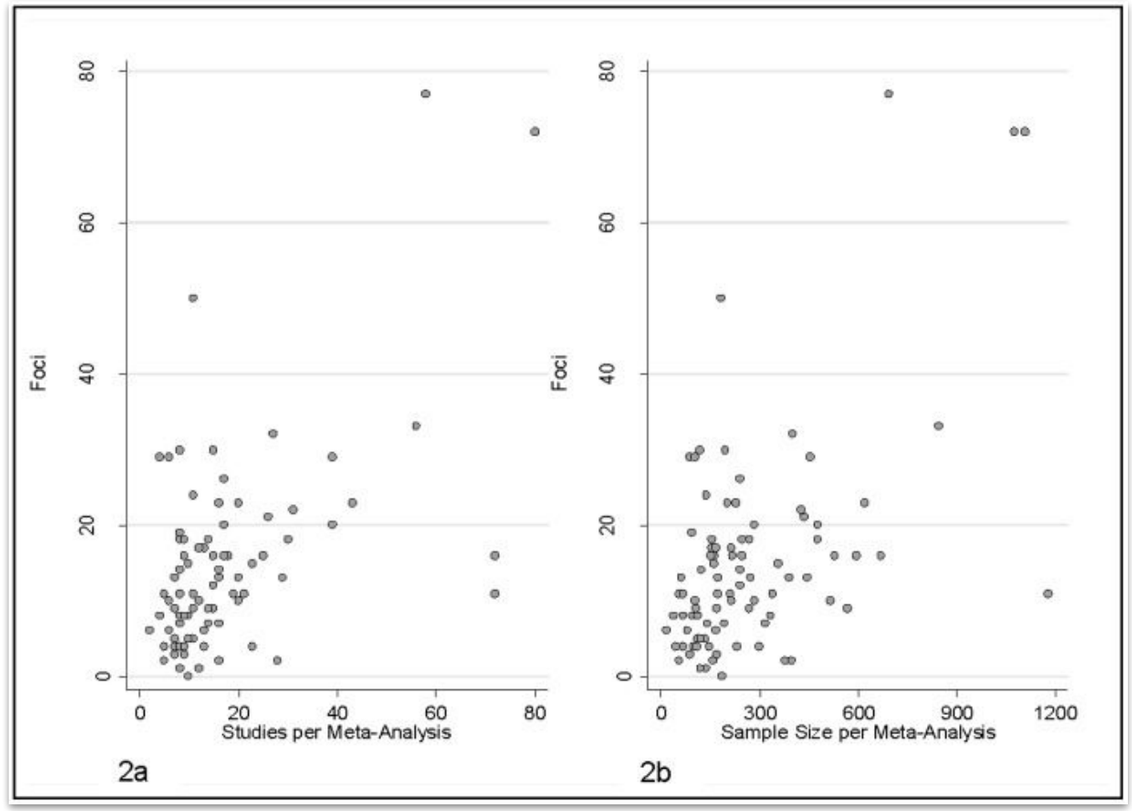


Image Source: David, Sean & Ware, Jennifer & Chu, Isabella & Loftus, Pooja & Fusar-Poli, Paolo & Radua, Joaquim & Munafò, Marcus & Ioannidis, John. (2013). Potential Reporting Bias in fMRI Studies of the Brain. PloS one. 8. e70104. 10.1371/journal.pone.0070104.

MyRBQM® Centralized Statistical Monitoring (CSM)

Fraud detection algorithms:

Correlation Matrix

Correlation matrix aims for **statistical dependence or association** between two variables discovering interrelations of variables for the **detection of fraud and tampering with data.**

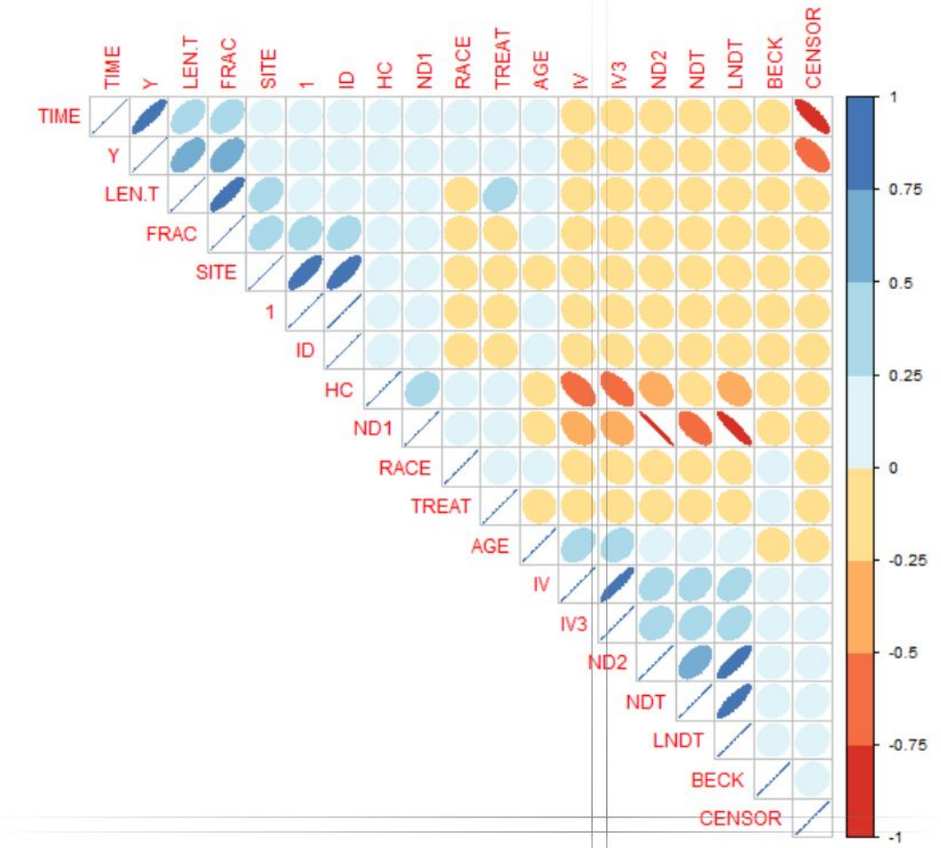


Image Source: MyRBQM® Portal

MyRBQM® Centralized Statistical Monitoring (CSM)

Fraud detection algorithms:

Histogram of all numeric values

A histogram is an accurate representation of the distribution of numerical data. It differs from a bar graph, in the sense that a bar graph relates two variables, but a histogram relates only one. It helps in **anomaly detection** in the process of **finding outliers in data sets**.

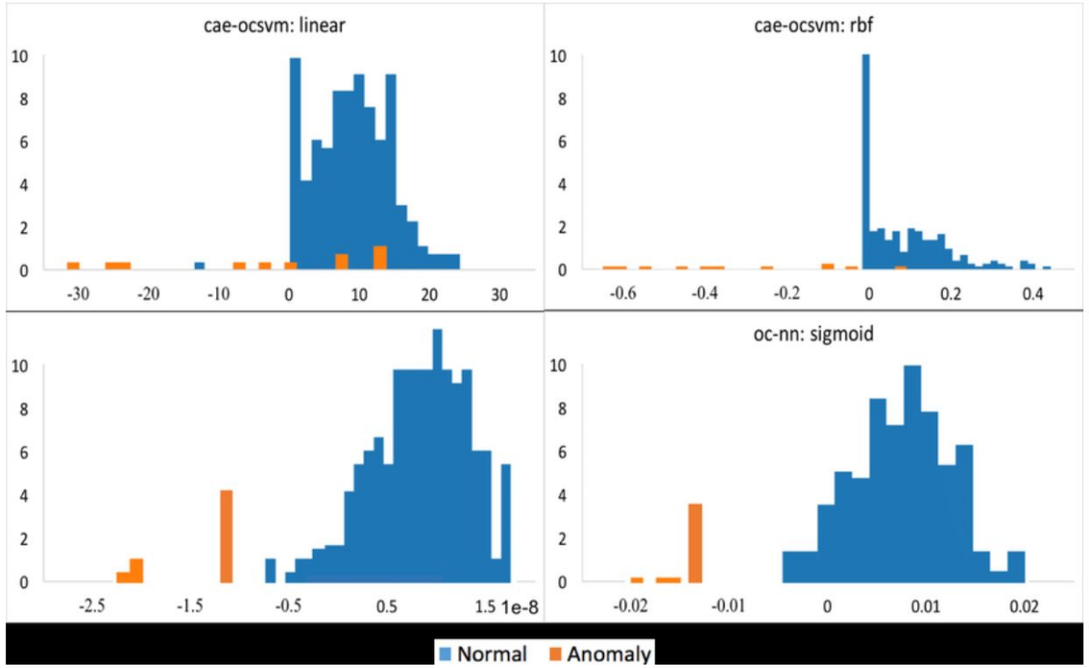
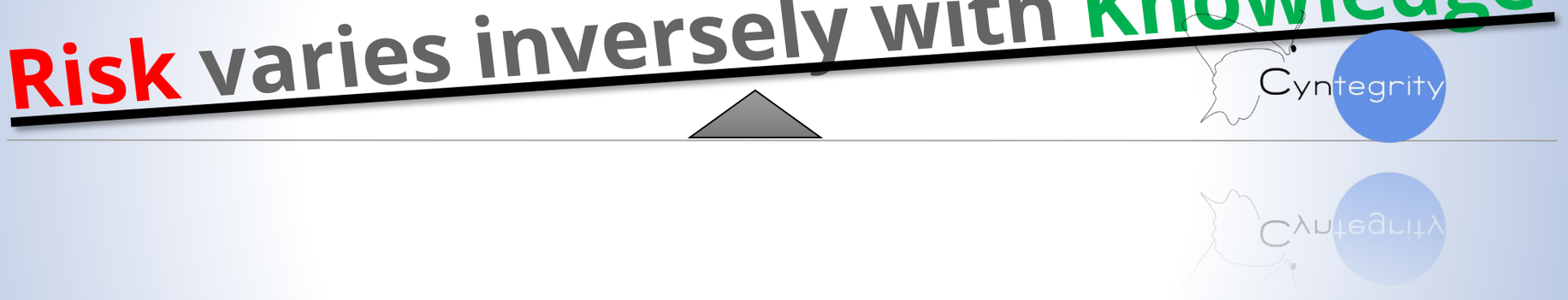


Image Source: Chalapathy et al. (2018) Anomaly Detection using One-Class Neural Network. arXiv:1802.06360v1

Risk varies inversely with Knowledge



Irving Fisher (1867-1947)

Q&A

Type in your questions using the chat box on the right side of your screen.

Artem ANDRIANOV
Chief Executive Officer
Cyntegrity Germany GmbH

Further Questions?

Contact information:

post@cyntegrity.com

artem.andrianov@cyntegrity.com

linda.bunschoten@cyntegrity.com

NEW: MyRBQM® Portal

Get Your Free User License Today!

www.cyntegrity.com

MyRBQM® Portal
powered by Cyntegrity

THANK YOU FOR LISTENING!