thoughtsphere

AI/ML DATA QUALITY & SMART CHECKS

Background

Risk Based Quality Management (RBQM) is the quintessential paradigm shift in how clinical trials are monitored. To ensure data quality in this new era of clinical research, data managers must supplement traditional field-level and subject-level data checks with AI/ML deployed data checks that look across the summation of data to identify data patterns, correlations, and deviations in real-time.

For many years SAS has been the go-to standard for implementing complex cross domain checks; however, this method which requires extensive upfront programming and off-line data transfers has quickly become antiquated and no longer meets the demands of the industry. Additionally, traditional coding platforms do not support AI/ML so they cannot dynamically adjust to the variety of data sources and the speed at which clinical data changes. **Now more than ever, data managers need a flexible solution that support the use of advanced AI/ML algorithms to explore aggregate data in conjunction with field and patient level checks to provide a 360° view of the data in real-time.**

ThoughtSphere Solution

To support the complex needs of data managers, ThoughtSphere leverages AI/ML to facilitate aggregate-level quality checks and provide a library of automated Smart Data Checks.

The platform uses unsupervised AI/ML algorithms and advanced analytic techniques to uncover atypical patterns, data signals, and outliers not detectable through traditional reviews. These quality checks are auto enabled for studies in the platform and start working behind the scenes to process ingested data and highlight data anomalies to end-users.

> The platform uses unsupervised AI/ML algorithms and advanced analytic techniques to uncover atypical patterns, data signals, and outliers not detectable through traditional reviews.

Examples of data quality checks leveraging the AI/ML models:

- Identify duplicate patients based on probability matching algorithms applied to key attributes such as DOB, gender, height, etc.
- Identify unexpected or unusual data correlations between specific data parameters such as AST vs. ALT, Heart Rate to Respiratory Rate, SYSPB vs. DIABP, etc.
- Uncover digit preference patterns across all numeric data points or a specific subset of values to identify procedural non-compliance and/or possible fraud
- Discover data outliers using multivariate distance and clustering algorithms
- Develop and deploy study-specific models to explore the data using the embedded Modeling and Analysis Programming (MAP) module

In addition to the advanced analytic checks available out-of-the-box, a library of pre-seeded data checks is also available to perform cross-variable reviews. Using our Smart Check technology, data check rules are automatically configured, irrespective of the study's metadata definition, using NLP and statistical-based text comparison algorithms. This enables data discrepancies to be identified, whenever new or changed data flows into the ClinHUB data lake, with no upfront programming or manual listing review required. Each data check in the library includes the associated query text so that queries can be directly pushed to EDC for site review & resolution.

Examples of data discrepancies surfaced automatically by Smart Data Check:

- AE treated with a ConMed, but no associated ConMed present
- Ongoing Medical History entry reported as an AE without a severity distinction (e.g., worsening, exacerbation, etc.)
- Duplicate ConMed entries with same or overlapping dates
- Duplicate AE entries with same or overlapping dates