



saama
#1 in AI Clinical Analytics

WHITE PAPER

AI-Based Clinical Data Review

Accelerate data cleaning and query generation with Smart Data Query, a transformative AI engine that brings state-of-the-art technology to your clinical data management practices.





Summary

High-quality clinical and scientific data—from electronic data capture (EDC) systems and third-party sources (labs, biomarkers, PK/PD, etc.)—is a prerequisite for accurate data analysis, reporting, and submission deliverables, as well as timely database lock.

Current approaches to ensuring data quality combine automated data checks and manual review processes to generate queries for further investigation. Manual review is incredibly labor intensive, time consuming, and error prone.

The vast majority of queries generated by data checks and reviews lead to no changes at all, or to changes to non-critical data points. A great deal of effort is expended to discover the less than three percent of all queries that require changes to data points that are critical to study success.

By training machine learning models on historical clinical data, with the goal of predicting data discrepancies and streamlining query management, clinical trial sponsors and CROs can significantly improve the manual data review process.

Accurate Data Review is Critical for Study Success

Clinical data review and analysis is essential for ensuring complete and clean study data, for accurate analysis and reporting on:

- Baseline characteristics (e.g., inclusion/exclusion criteria violations)
- Randomization/stratification (e.g., multiple enrollments for the same patient)
- Compliance (e.g., missing study visits)
- Disposition (e.g., discontinuation reasons)
- Efficacy and safety (e.g., data outliers)

To identify discrepancies and generate queries, data management teams at sponsor organizations and CROs use a combination of automated data quality (DQ) checks and manual reviews.

Automated DQ checks involve the monitoring and maintenance of hierarchical libraries: Global, Therapeutic Area, Indication, Project, etc. New studies inherit metadata-based DQ checks from previous studies, and custom DQ checks can be added based on any metadata that's specific to the new studies.

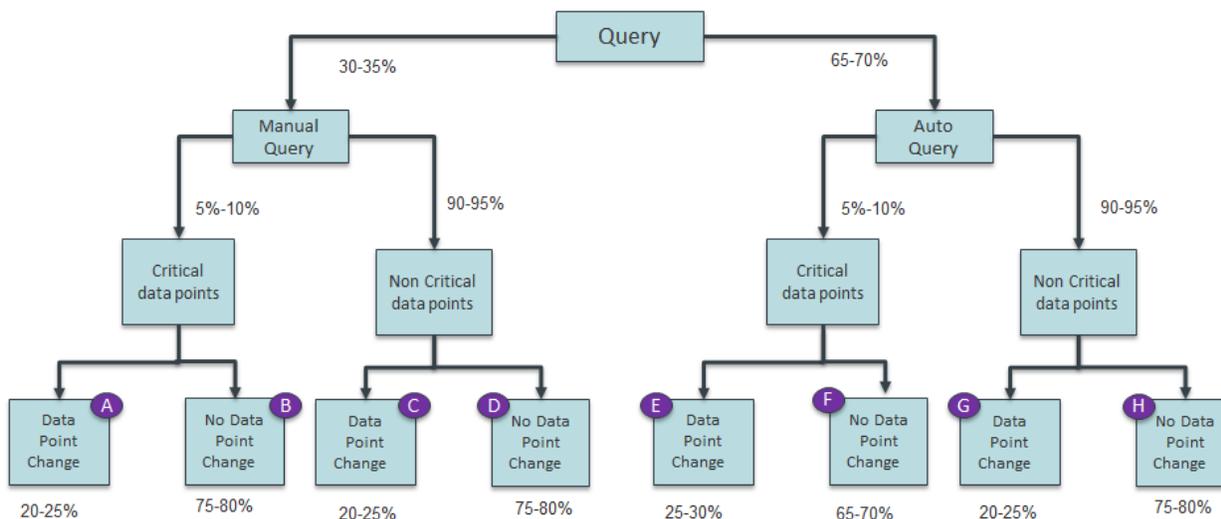
Manual review typically employs a package of programming and visualization tools, which requires significant effort to build for each study. Significant effort is also required to generate data listings and look them over to discover issues, inconsistencies, and anomalies.

The Query Landscape: Finding a Needle in a Haystack

A typical clinical study involves 30–50 raw CRF datasets with 40–60 variables each, and subject data across multiple visits can result in millions of data points.

Queries based on these data points can be classified by how they're generated, their association to critical or non-critical data, and whether or not they lead to data point changes. Queries that lead to changes in critical data points are, obviously, the most important.

An analysis of data from different sponsors confirms that these high-value queries are quite rare, amounting to less than three percent of all queries. In the chart below, category A represents high-value queries from manual review and category E represents high-value queries from automated DQ checks.



AI Can Identify Discrepancies Faster and Improve Query Management

Machine learning models, trained on historical clinical trial data, can identify potential data discrepancies much faster than manual reviews can.

This AI technology is made possible by the ability to ingest data from all your source systems, bring it together in a unified study data model, and update it in real time. This is the foundational capability of Saama's **Life Science Analytics Cloud (LSAC)**.

Saama's machine learning data review solution, **Smart Data Query (SDQ)**, gives you all the components you need to identify data discrepancies, generate query text, and manage queries. This domain-centric solution integrates bi-directionally with your EDC and other source systems and delivers a unified experience for query tracking, data review, and workflow.

By learning patterns from historical data, the machine learning model decides whether or not queries need to be raised. Using SDQ, data management teams can:

- Classify predictions across 10+ categories and 80+ subcategories
- Generate query text automatically
- Review only data points with clinical discrepancies
- View clinical data in SDQ with deep integration with your EDC
- Track and measure machine feedback
- Approve, reject, and put discrepancies on hold

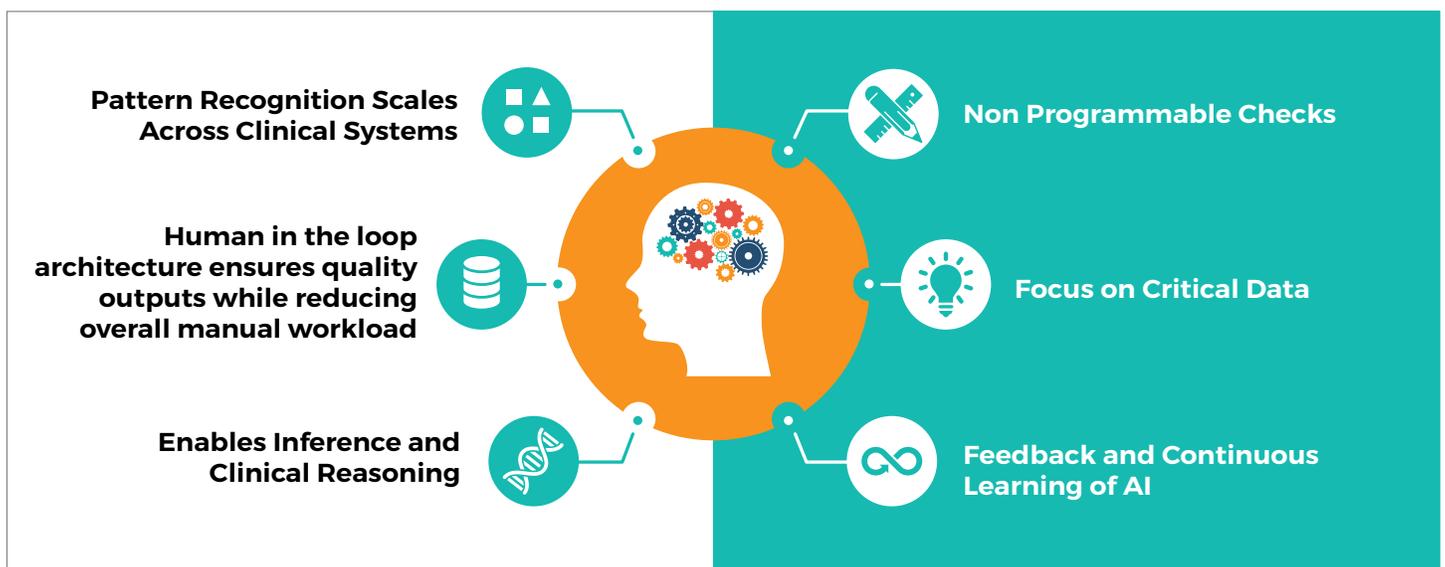
Before SDQ

- Need to review each and every data point
- Manual tracking and reviewing of every data discrepancy required
- Manual review/inspection system
- Spreadsheet (Jreview/SAS) based systems used to review exports from EDC with data refresh
- Reliance on data from third-party providers (e.g., CROs) for raising queries

After SDQ

- AI generated to-do lists and clinical queries
- Review only data points with clinical discrepancies
- Machine-based feedback mechanism (human in the loop) with AI traceability
- Easy SSO (OAuth and SAML) with centralized, smart to-do lists for tracking and downloads
- Auto generated discrepancies and explanations; ability to use APIs to import queries to EDCs

Manage Data More Effectively with SDQ



Smart Data Query (SDQ) Solution Components

Data Acquisition

- Adapters for pulling in data and metadata in near real time from EDC and CTMS
- Data configurations for file-level and metadata checks
- Intelligent file watcher utility for pulling in third-party files for metadata checks

Model Management

- Pipeline approach to model training, testing, performance tuning, and baseline deployment
- Scalable model management for additional therapeutic areas and indications
- Extendable with other EDC data structures (based on primary and auxiliary data points)
- Re-usable pipeline for model re-training

Interactive User Interface

User Interface for Model selection, viewing the prediction details for question and filtering the data

- Enable users to accept, override system recommendations and to view review history
- User Interface to continuously view Model performance history to decide on promotion and re-training
- Simplified Interface for both Verification (can be used as a GxP validation step) and Production

Systems Integration

- Unified experience for query tracking, data review (manual, auto queries) and review workflow
- Bi-directional EDC integration for Query actions from Review User Interface
- Integration with Machine learning based standardization, analysis and submission deliverables
- Integration with analysis, visualization and statistical tools based on fine grained access.

Machine Learning Model Development and Improvement

One of the most exciting advantages of SDQ is its ability to get better and better over time. In the beginning, the base machine learning model is created through multiple teaching cycles on your historical data. After that, the model uses ongoing study data to continuously learn and improve.

This virtuous circle of enhancement requires a “human in the middle” to provide feedback, which the model uses to become more intelligent. The framework below shows the specific processes for prediction review, feedback storage, model retraining, and redeployment.

Conclusion

Machine learning models trained on CRF data can help clinical data management teams, medical directors, and medical monitors identify data quality issues and manage queries in a more efficient manner.

Saama’s AI-powered engine—SDQ—automates and accelerates your data management processes, so you can instantly answer questions like these:

- Is a concomitant medication consistent with an AE term?
- Are duplicate medications given for the same condition?
- Are related AEs, such as RECURRENT FEVER and UNKNOWN FEVER, of the same toxicity?
- Are AE terms like DENTAL EXTRACTION linked to non-drug treatments?

To learn more about SDQ and arrange a demo, contact Saama at info@saama.com or 408-371-1900, or visit us online at saama.com.



About Saama Technologies, Inc.

Saama is the #1 AI clinical analytics platform company, enabling the life sciences industry to conduct faster and safer clinical development and regulatory programs. Today, 50 biotech companies use Saama's award-winning Life Science Analytics Cloud (LSAC) platform on over 1,500 studies, including many of the top 20 pharmaceutical companies. LSAC's rich applications facilitate an unprecedented, authoritative oversight of comprehensive clinical research data, enabling companies to file New Drug Applications (NDAs) more efficiently and bring drugs to market faster. Discover more at www.saama.com and follow Saama @SaamaTechInc



sales@saama.com
www.saama.com
+1 888.205.3500

Saama and other marks used herein are trademarks of Saama Technologies, Inc.
Copyright © 2020 Saama Technologies, Inc. 07/20