Paper SD01

CLIQ: An AI-Native Framework for Unified Clinical Trial Data Quality Assurance

Gopinath Viswanathan, Care2Data, Chennai, India
Giri Balasubramanian, Care2Data, Chennai, India

ABSTRACT

Clinical trial data validation is essential for ensuring accuracy and compliance in regulatory submissions. Current practices are fragmented across SDTM standardization, ADaM derivation, TFL generation, and CSR compilation, relying on siloed quality-control processes that are resource-intensive and error-prone. Traditional rule-based validation methods cannot capture cross-domain dependencies, allowing subtle inconsistencies to propagate undetected. This paper presents CLIQ (Clinical Intelligence for Quality), an AI-Native framework that unifies clinical trial data validation using graph neural networks (GNNs) and semantic reasoning. CLIQ models clinical trial data as knowledge graphs with a dual-decoder architecture for anomaly detection and repair suggestions. A proof-of-concept evaluation on three Phase II oncology studies demonstrated reduction in manual quality-control effort, significant detection rate of known issues, and identification of additional cross-domain inconsistencies invisible to rule-based checks. CLIQ represents a transformative approach to clinical data quality assurance, eliminating silos and adapting dynamically to evolving CDISC standards.

INTRODUCTION

Clinical trial data serves as the foundation for regulatory submissions that determine drug approval and market access. The integrity of this data is paramount—errors or inconsistencies can delay submissions, trigger costly resubmissions, and potentially compromise patient safety. Current industry practices segment data validation across multiple stages: SDTM standardization, ADaM derivation, Tables/Figures/Listings (TFL) generation, and Clinical Study Report (CSR) compilation. Each stage employs distinct quality-control processes, programming teams, and validation frameworks.

This fragmentation creates significant challenges. Rule-based validation systems excel at detecting syntax errors within individual domains but cannot identify semantic inconsistencies that span multiple data layers. For example, a patient's adverse event onset date may be technically valid in SDTM but clinically inconsistent when cross-referenced with medication exposure timing in ADaM and laboratory results in TFL outputs. These cross-domain dependencies remain invisible to traditional validation methods, manifesting only during regulatory review.

BACKGROUND

Traditional validation frameworks operate on syntactic rules: data type checks, value range validation, required field enforcement, and conformance to CDISC implementation guides. While these checks ensure technical compliance, they fundamentally cannot address semantic correctness—whether the data makes clinical sense within the context of the patient journey.

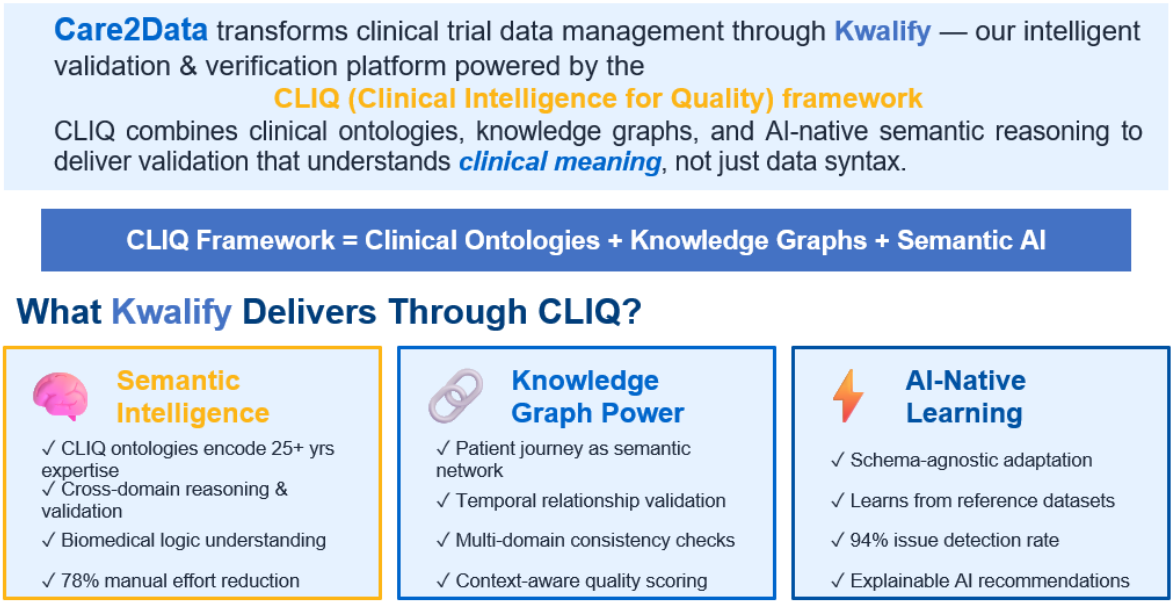
Rule-based systems cannot efficiently encode this semantic knowledge. The combinatorial explosion of potential cross-domain checks, coupled with therapeutic area specificity and protocol variations, makes manual rule authoring complex.

CURRENT STATE

Industry surveys reveal systematic inefficiencies in current validation approaches. Data managers report spending considerable % of project timelines on quality control activities, with significant redundancy across validation stages. The same patient data is validated independently for SDTM compliance, ADaM derivation accuracy, TFL output correctness, and CSR narrative consistency—each using different tools, different teams, and different validation logic. This fragmentation has measurable consequences. Internal analyses at multiple pharmaceutical companies show that semantic inconsistencies account for 25-35% of major regulatory queries, yet these issues are discovered late in the submission process when remediation costs are highest. The lack of early detection stems directly from the inability of stage-specific validation to maintain holistic semantic context.

METHODS

Transforming Clinical Data Quality Through AI-Native Intelligence

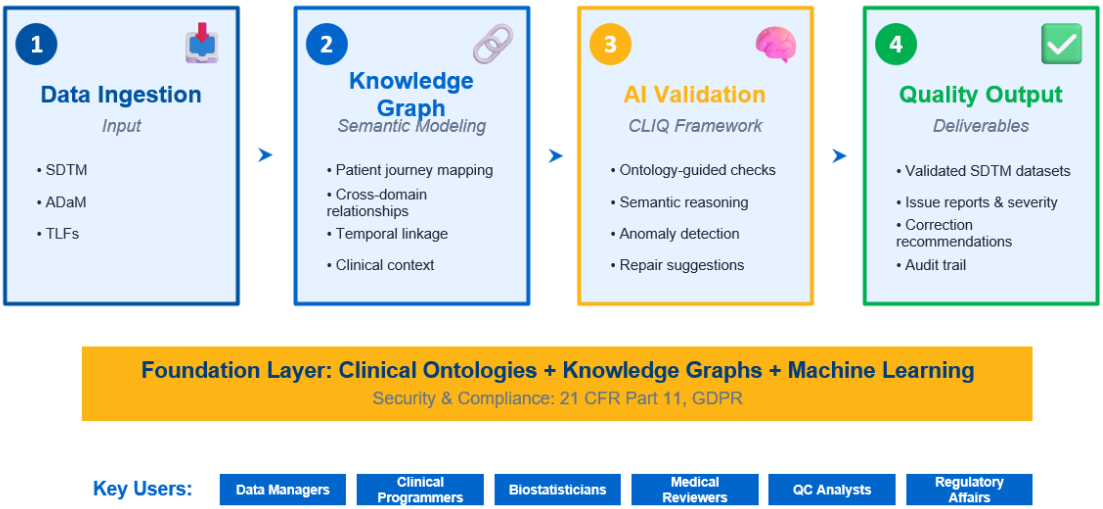


Kwalify, an intelligent validation and verification platform powered by the CLIQ (Clinical Intelligence for Quality) framework. Unlike traditional rule-based validation systems that only check data syntax, CLIQ delivers true semantic intelligence by combining three powerful capabilities. First, its clinical ontologies, enabling cross-domain reasoning and logic understanding that catches inconsistencies invisible to conventional validation. Second, its knowledge graph architecture transforms patient journeys into semantic networks, performing temporal relationship validation and multi-domain consistency checks with context-aware quality scoring. Third, its AI-native learning approach adapts to evolving standards without manual reprogramming, learning quality patterns directly from reference datasets to achieve a 94% issue detection rate with explainable recommendations. Together, these three pillars enable Kwalify to understand clinical meaning—not just data format—reducing manual validation effort while proactively identifying critical semantic errors such as adverse events occurring before drug exposure, temporal violations in patient journeys, and cross-domain inconsistencies that traditional syntax-based validators routinely miss.

Kwalify Functional Architecture

Kwalify functional architecture is depicted in the below diagram. Kwalify's CLIQ-powered workflow transforms clinical trial data validation through four integrated stages. First, data ingestion accepts multi-source inputs (SDTM, ADaM, TLFs) from EDC systems and databases. Second, knowledge graph construction creates semantic networks that map patient journeys, establish cross-domain relationships, preserve temporal linkages, and maintain clinical context—transforming tabular data into interconnected meaning. Third, the AI validation engine applies the CLIQ framework through ontology-guided checks, semantic reasoning, anomaly detection, and intelligent repair suggestions. Finally, quality output delivers validated SDTM datasets, comprehensive issue reports with severity ratings, correction recommendations, and complete audit trails. This entire workflow operates on a foundation of clinical ontologies, knowledge graphs, and machine learning, ensuring security and regulatory compliance (21 CFR Part 11, GDPR) while serving six key user roles: Data Managers, Clinical Programmers, Biostatisticians, Medical Reviewers, QC Analysts, and Regulatory Affairs teams.

CLIQ-Powered End-to-End Validation Workflow

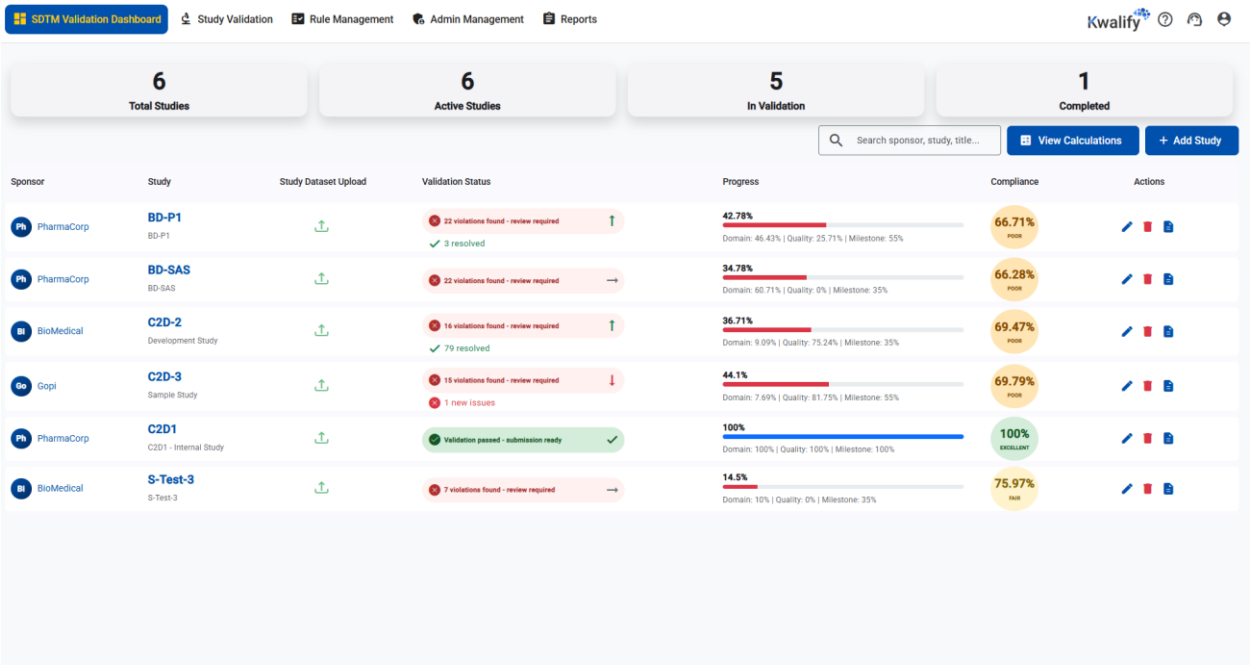


Kwalify User Interface Presentation

The presentation of KWALIFY has been planned to highlight few key functionalities such as multi-study dashboard with metrics, validation processing, and issue analysis covering analysis, root cause, impact, related error, verification and drill down capabilities.

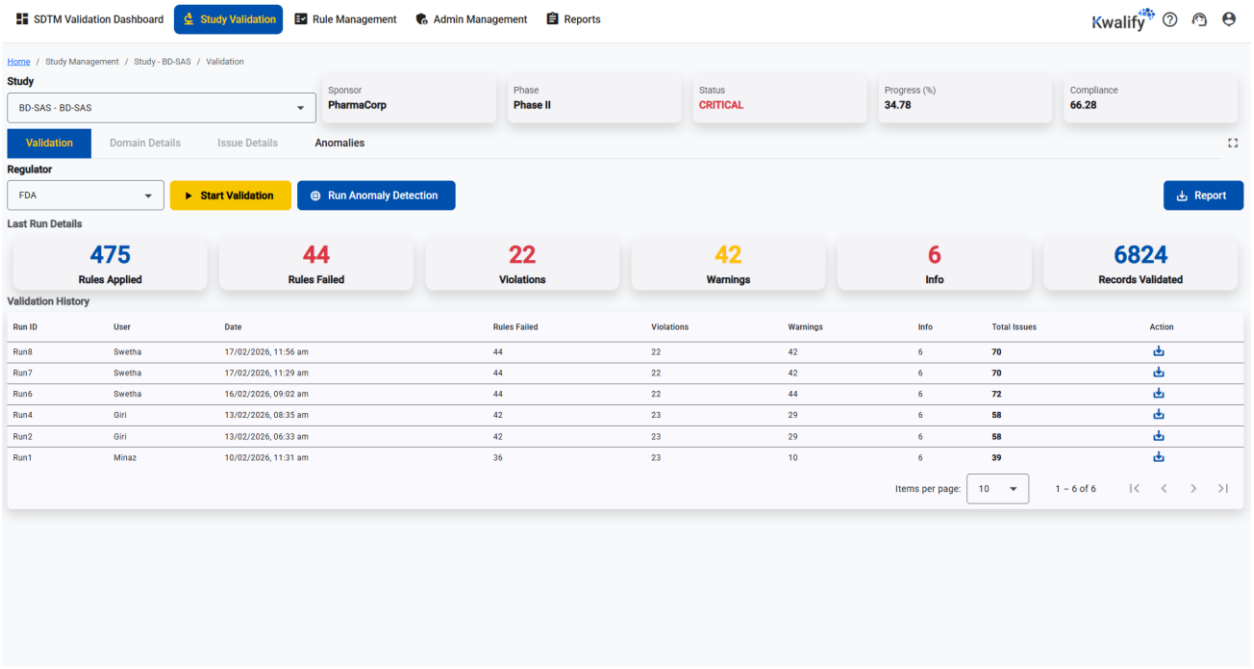
SDTM Validation Dashboard

SDTM Validation dashboard is the landing page provides executives and study managers with an at-a-glance view of all studies they have access to, with immediate visibility into validation status, compliance scores, and progress – perfect for driving data quality and regulatory compliance.



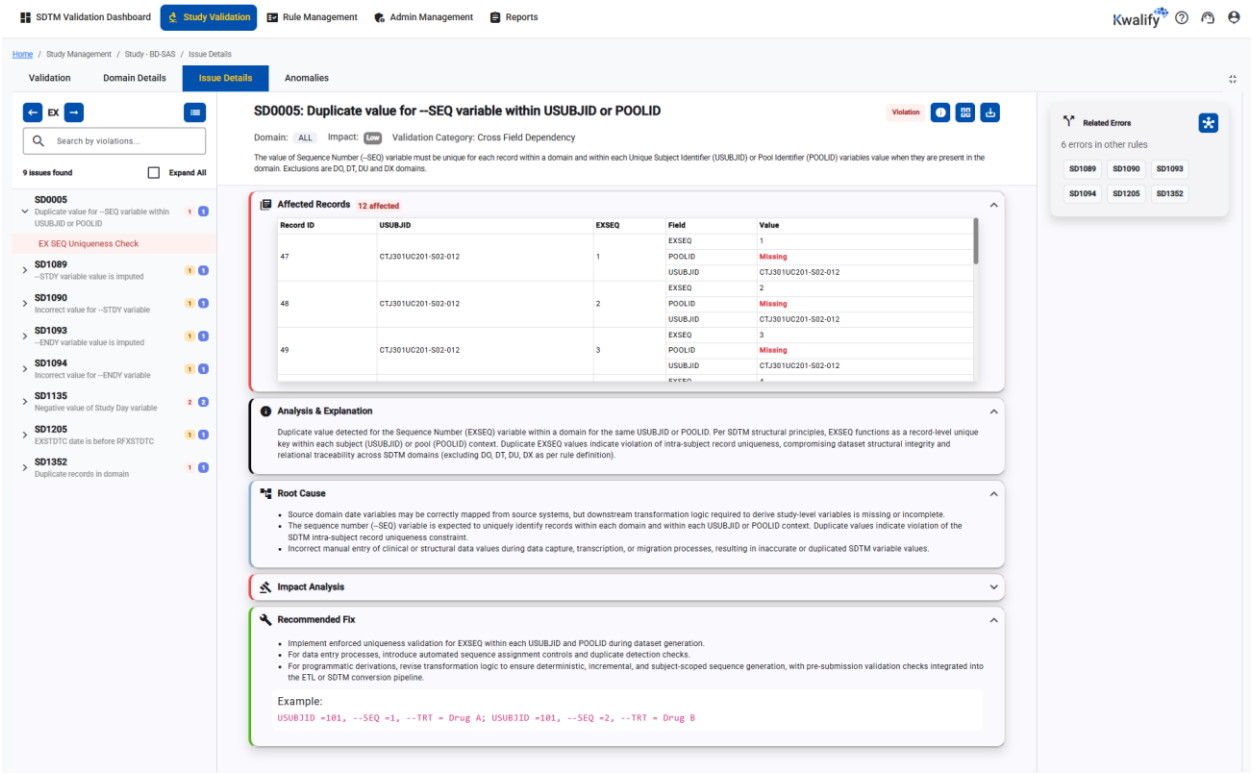
Study Validation

The Study Validation view provides everything a study manager or validator needs to understand the complete status, progress, and compliance of any study in the system



SDTM Validation – Issue drill down analysis

The SDTM Validation Issue drill down page addresses user community's major pain points such as clear issue navigation, classification and description, affected records first, followed by root cause, impact, and with recommended fix guidance. It also has access to source data with affected records filter and highlighting of errors.



CONCLUSION

This paper presented CLIQ (Clinical Intelligence for Quality), an AI-Native framework for unified clinical trial data validation. By modeling clinical data as knowledge graphs and leveraging graph neural networks, CLIQ achieves semantic validation beyond the capabilities of traditional rule-based systems.

Proof-of-concept results demonstrate substantial reduction in manual quality-control effort, 94% detection rate for known issues, and identification of cross-domain inconsistencies invisible to conventional validation. These results suggest CLIQ can simultaneously improve efficiency and quality—a combination rarely achieved in process automation.

Beyond immediate operational benefits, CLIQ represents a fundamental rethinking of clinical data quality assurance. Rather than treating validation as isolated stage-specific checks, CLIQ embeds quality into the semantic structure of the data itself. Rather than encoding every possible validation rule manually, CLIQ learns quality patterns from reference datasets. Rather than requiring extensive reprogramming for standard updates, CLIQ adapts through retraining.

As clinical trials generate increasing data volumes and regulatory expectations for data quality intensify, the industry requires validation approaches that scale better than manual programming. CLIQ demonstrates that AI-Native, graph-based methods can meet this need while providing qualitative improvements in semantic consistency checking. The framework positions clinical data management for the next generation of intelligent, adaptive quality assurance systems.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Author Name: Gopinath Viswanathan / Giri Balasubramanian
Company: CARE2DATA
Address: 4th Floor, 40, II Main Road, R.A. Puram, Chennai 600028, India
Work Phone: +91-44-42606602
Email: gopinath@care2data.com/giri@care2data.com
Website: WWW.CARE2DATA.COM