Paper SD08

Cybernated Validation of Reports

Giri Balasubramanian, CARE2DATA, Chennai, India

ABSTRACT

Clinical Summary Reports generated post analysis activity undergoes another round of validation in the name of double programming in order to confirm the outputs for final submission with the sponsors (Pharma) and later with regulatory bodies.

The purpose of this presentation is to present a semantic intelligence powered platform to automate the validation and verification of clinical reports such that the effort expended in manual verification of the clinical reports by CROs/Pharmaceuticals can largely be reduced

Clinical Report Validation is a domain intensive space and applications intending to automate this process need to deal with huge amounts of unstructured and structed textual data. The inter-dependencies within summary tables, cross table checks and associations with the underlying listings data might be very intuitive to the human eye, however not so apparent to a machine. Semantic web technology enables weaving the context of the data into a web of interconnected information that is very useful to determine the subtle relationships across the summary and listing files with minimal manual intervention.

The validation and verification process feeds off the semantic layer to automatically detect which columns from the listings files need to be summarized for certain sections of the summary tables.

The interconnected nature of the study metadata such as the sponsoring organization, Therapeutics Category, the Study Titles and the summaries that are part of a particular study provide valuable information on the common vocabularies and their hierarchies. This information is best handled in the form of knowledge graphs as they can provide a single snapshot of integrated information across the entities.

Finally, Semantic Web Data Framework uses RDF standards to define and store the knowledge graph which is FAIR compliant.

INTRODUCTION

Clinical reports generated by experiments, research studies, drug launches and clinical trials to name a few needs to be validated before formal submission of the reports and regulatory approvals and undergoes independent round of validation in the name of double programming in order to confirm the outputs for final submission with the sponsors (Pharmaceuticals) and later with regulatory bodies.

Independent programmer creates a separate program to generate the required results working without reference to the past program and without interaction from the main programmer to avoid any bias. The results of the two programs are compared and if any noticeable unexplained or unacceptable differences result in failing validation and quality check. This approach has become gold standard for the validation of Clinical Trial Reports which comprises of Analysis datasets, TLFs including associated programs.

Adoption of Independent Programming brings in unbiased quality check, ensuring all the output results are validated and QC' ed without much interaction with original programmer, thereby reducing the risk of errors and increase quality of deliverables adhering to 21CFR Part 11 guidance.

Current process of Independent Double programming has its own challenges and limitations. Being a continuous activity during the trials towards submissions by a CRO with Pharma, it is lot more of a human challenge to maintain consistency in quality throughout the life cycle of a trial and final submission with regulatory in reducing the risk of errors.

Many solutions have been proposed and implemented by Pharmaceutical and CROs in the last decade in order to address the challenges and limitations identified in the existing model of validation and quality check including risk-based models and automated comparison of results using macros.

Each of the advancement done so far in the process of validation and verification has its own limitations when it comes to usage by the experienced Statistical Programmers and Statisticians.

With recent advances in the deployment and usage of semantics, CARE2DATA's KWALIFY – intelligent data driven platform was able to validate & verify the consistency in the numbers reported across various tables, listings in an intelligent manner which was largely manual in the past, carried out by the source organizations (i.e. the Bio-Tech / Pharma / Research Organization that generated the report) and sometimes outsourced to the Clinical Research Organization(s) (CROs) for validating the reports.

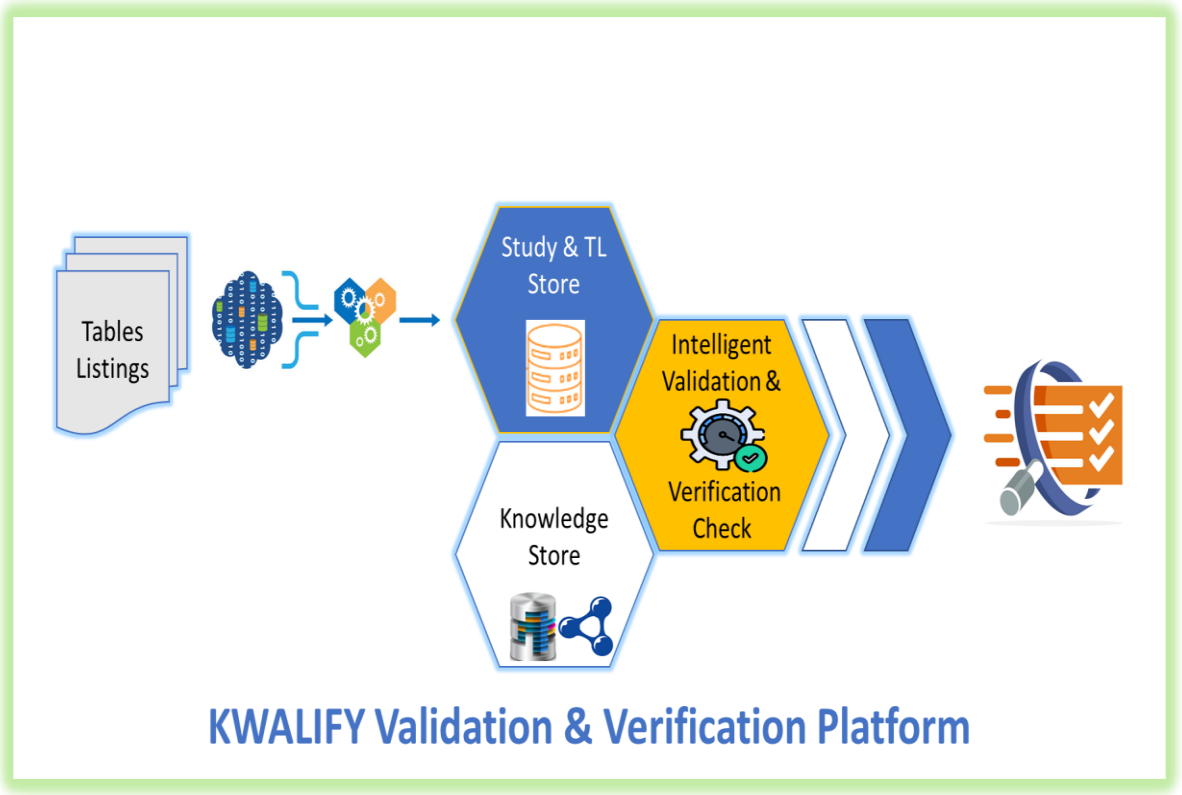
Semantics organizes data into well-defined categories with clearly defined relationships. Classifying information in this way enables humans and machines to read, understand and infer knowledge based on its classification. CARE2DATA's KWALIFY uses this underlying concept in graph model to build intelligent relationship using their ontology to compare and cross verify the results and permits the user to confirm the findings on the false positive and false negative items under the discrepancy truth table.

KWALIFY: *'the solution'*

KWALIFY is a data-driven intelligent validation platform that enables CROs, Pharmaceuticals to accelerate their journey to resilient, self-healing validation across their entire IT estate. The platform empowers Senior Clinical Statistician teams to detect and resolve issues quickly, and automatically predict & prevent future problems before they happen.

It performs the role of Validation & Verification (V & V) thereby bringing consistent quality and dependability of clinical trial analysis results. KWALIFY is put to use throughout the life cycle of trial prior to any interim or final submissions being made by CROs or Pharmaceuticals with regulatory body. In addition, KWALIFY facilitates availability of experienced statisticians to be available for more than one study or trial maintaining consistent quality throughout the data flow cycle of a trial.

Functional Architecture of KWALIFY is depicted in the below diagram. The diagram highlights the flow of the trial data from reports, which is extracted and stored in databases. Semantics is built around the extracted information in knowledge graph model which facilitates the process of intelligent validation & verification check using CARE2DATA's proprietary V & V framework. KWALIFY publish the analysis results for the Sr Statisticians and Statistical programmers to confirm the checks as part of the process flow.



The presentation of KWALIFY has been planned to highlight few key rules of validation and verification that is being performed at a Table level with and without using Listings as a reference including Cross Table validation.

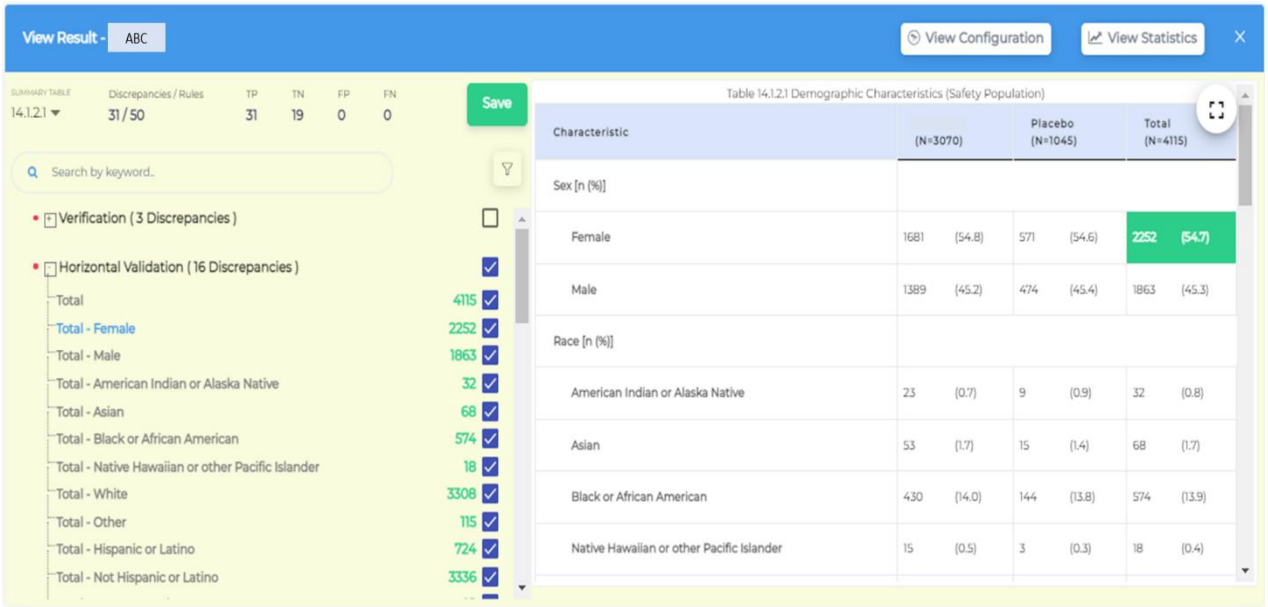
Below sections explains the demonstration of the Table level validation that is being performed by KWALIFY.

Table Validation without Listing Reference

Focus towards this demonstration is to highlight how the validation is performed on the Total Column of a Table automatically by KWALIFY. Here, the system is able to validate the sum and present the results for any discrepancy during the summation type of validation. This type of validation is called Horizontal Validation as per the system.

Table 14.1.2.1 Demographic Characteristics
(Safety Population)

Characteristic	(N=3070)	Placebo (N=1045)	Total (N=4115)
Sex [n (%)]			
Female	1681 (54.8)	571 (54.6)	2252 (54.7)
Male	1389 (45.2)	474 (45.4)	1863 (45.3)
Race [n (%)]			
American Indian or Alaska Native	23 (0.7)	9 (0.9)	32 (0.8)
Asian	53 (1.7)	15 (1.4)	68 (1.7)
Black or African American	430 (14.0)	144 (13.8)	574 (13.9)
Native Hawaiian or other Pacific Islander	15 (0.5)	3 (0.3)	18 (0.4)
White	2466 (80.3)	842 (80.6)	3308 (80.4)
Other	83 (2.7)	32 (3.1)	115 (2.8)
Ethnicity [n (%)]			
Hispanic or Latino	518 (16.9)	206 (19.7)	724 (17.6)
Not Hispanic or Latino	2512 (81.8)	824 (78.9)	3336 (81.1)
Not Reported	35 (1.1)	13 (1.2)	48 (1.2)
Unknown	5 (0.2)	2 (0.2)	7 (0.2)
Age (years)			
n	3070	1045	4115
Mean (std)	45.3 (15.54)	44.4 (15.24)	45.0 (15.47)
Median	46.0	45.0	45.0
Q1, Q3	32.0, 58.0	32.0, 56.0	32.0, 57.0
Min, Max	18, 94	18, 85	18, 94



Below section demonstrates automated vertical roll up based validation which is performed on the group, sub-group or sub section level within a Table by KWALIFY. Here, the system is able to do roll-up sum of a sub-group and validate it against the group total and present the results for any discrepancy as part of validation. This type of validation is called Vertical Validation as per the system.

14.3.2.2
Adverse Events by System Organ Class and Preferred Term
(Safety Set)

Sytem Organ Class Preferred Term	Treatment A (N=40) n (%)	Treatment B (N=159) n (%)
Any Adverse Event	12 (30.0)	41 (25.8)
Nervous system disorders	7 (17.5)	32 (20.1)
Somnolence	4 (10.0)	20 (12.6)
Dysgeusia	2 (5.0)	9 (5.7)
Headache	1 (2.5)	1 (0.6)
Dizziness	0	8 (5.0)
Dysarthria	0	1 (0.6)
Hypersomnia	0	1 (0.6)
Syncope	0	1 (0.6)
Respiratory, thoracic and mediastinal disorders	4 (10.0)	5 (3.1)
Cough	2 (5.0)	3 (1.9)
Dry throat	1 (2.5)	1 (0.6)
Oropharyngeal pain	1 (2.5)	0
Throat irritation	1 (2.5)	1 (0.6)
Psychiatric disorders	2 (5.0)	3 (1.9)

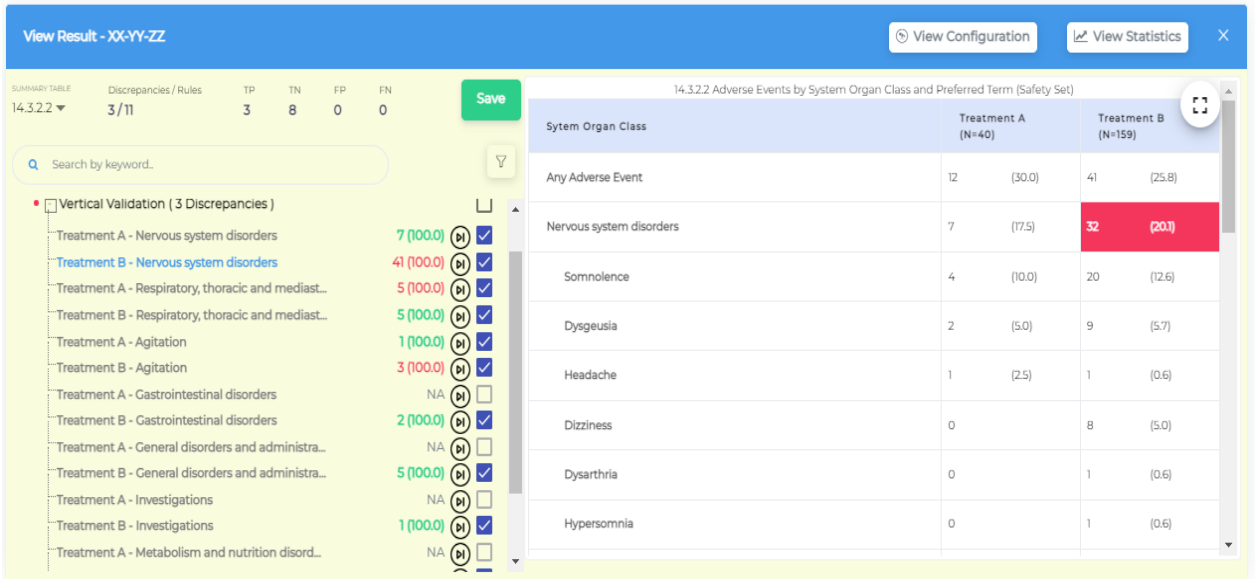
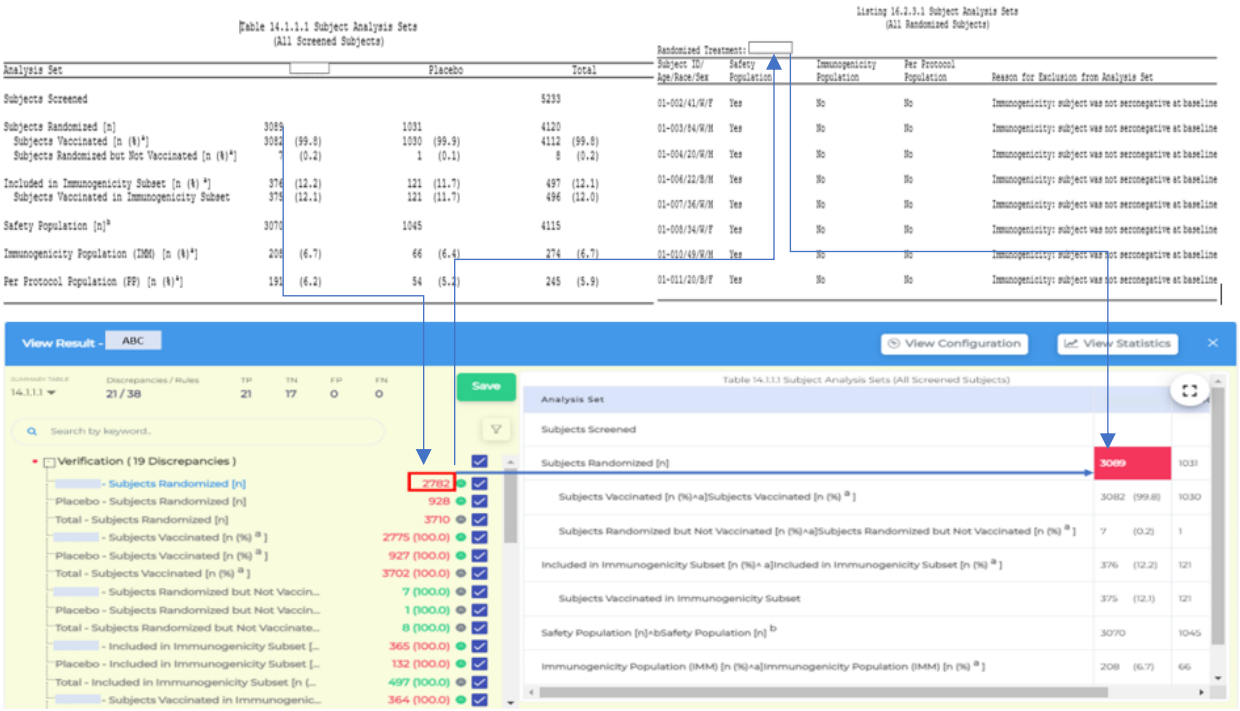


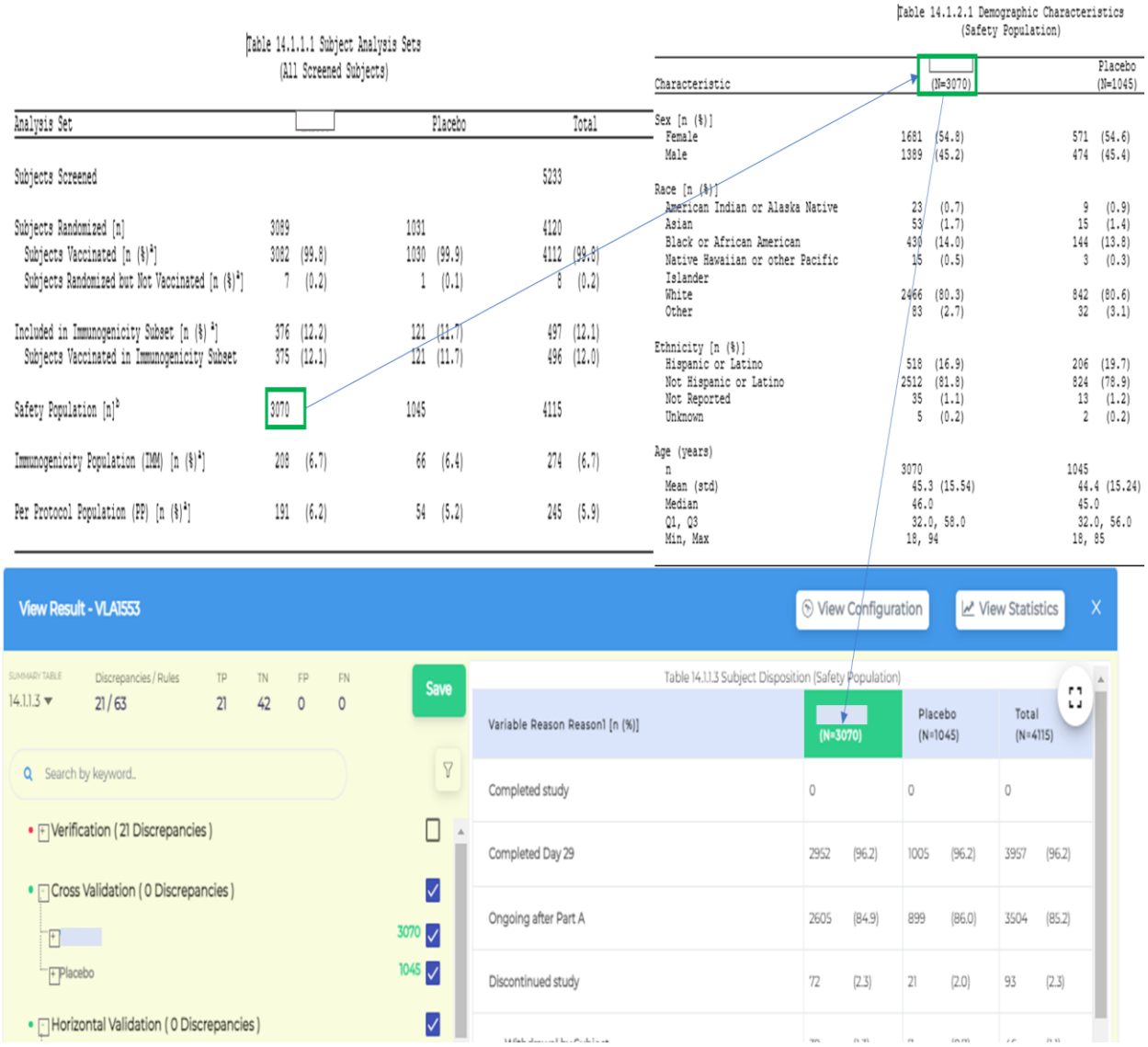
Table Validation using one or more Listing Reference

This section demonstrates the capability of using Listing as a reference to validate each of the data points in the generated Table output. Identified discrepancy is auto verified by the system using data triplets and is one of key features of KWALIFY using Semantic Web Technology. In this example below, the discrepancies flagged by the system are not false positives, and it helps in reducing the manual effort expended by the user in verifying the flagged discrepancies raised by the system.



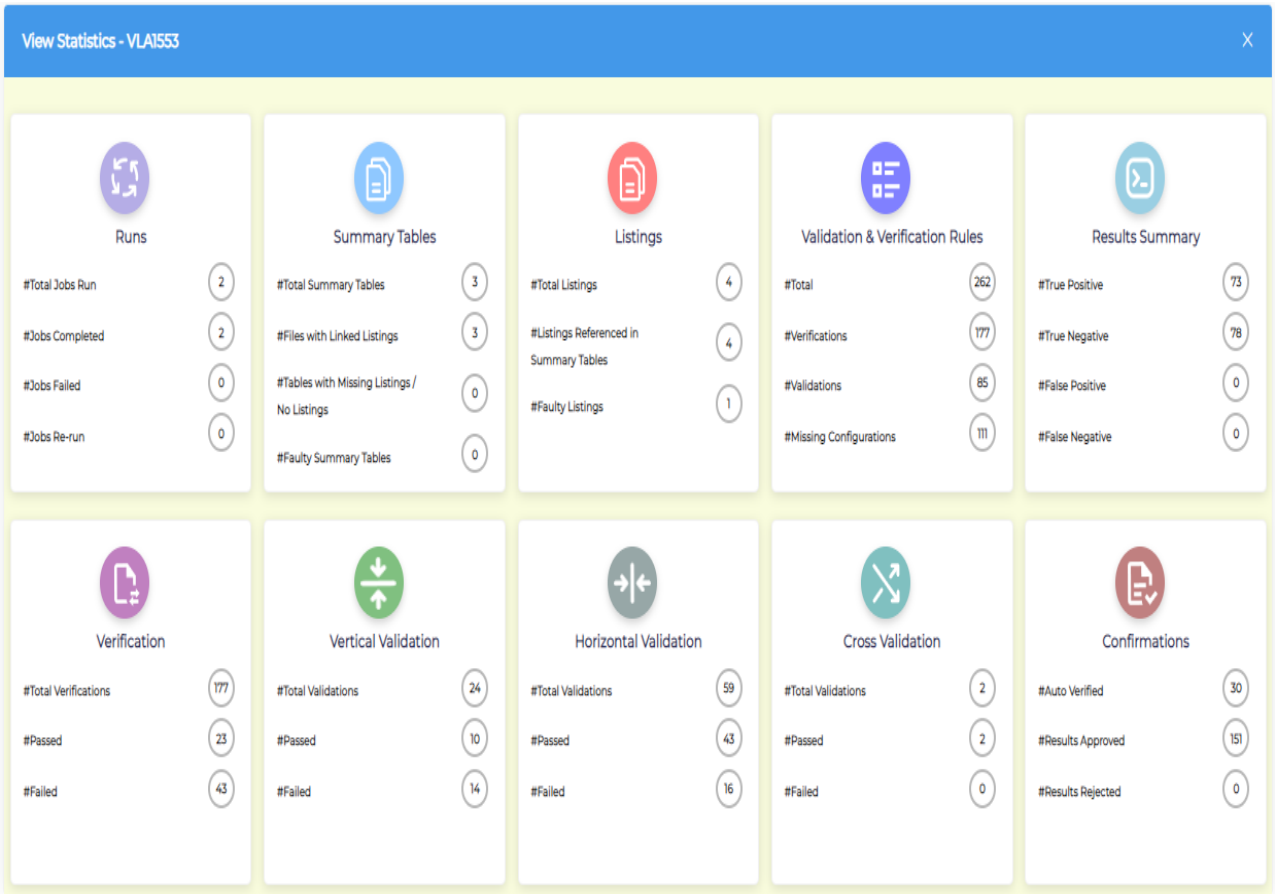
Cross Table Validation

In addition to vertical, horizontal and using listing reference, KWALIFY has a feature to validate across table outputs. By this way, quality of table outputs is cross verified and consistency is maintained throughout the report components and can be extended for the entire data flow in a trial. This demonstrates the use of data context which helps to map across tables and generate results for the user to confirm.



Result Analysis

KWALIFY finally, generate the summary of results for the user who tries to validate a set of TLs.



BENEFITS

KWALIFY brings in innumerable number of benefits to the CROs and Pharmaceuticals towards V & V activity that is needed prior to submission of reports. Few of the key benefits has been highlighted below:

- Consistent Quality
- Reduces the risk of errors
- Expedites Validation Cycle, and thereby Contributes Time to Market
- Handles data volume with variety of data
- Reuse V & V rules across similar studies
- Intelligent data mapping using data triplets
- Unbiased QC
- Ability to handle validation throughout the data flow in a trial

CONCLUSION

KWALIFY demonstrates the capability of performing automated validation of TLs using data triplets stored in knowledge graph. This approach leads to intelligent data driven V & V model to ensure consistent quality of deliverables, thereby reducing the risk of errors and enabling the Quality Control throughout the life cycle of data flow in a trial.

This paves way for the industry to focus on intelligent data driven models to handle large volumes of data / files without compromising on quality and with less manual effort.

REFERENCES

A. What makes a Good QC Programmer?
https://www.pharmasug.org/proceedings/2018/LD/PharmaSUG-2018-LD12.pdf
B. An Ontology to enable semantic BIM based data quality assurance and quality control
https://conferences.lib.unb.ca/index.php/tcrc/article/view/612/84
C. Baffling Concept of True Positive and True Negative
https://towardsdatascience.com/baffling-concept-of-true-positive-and-true-negative-bffbc340f107

CONTACT INFORMATION

Your comments and questions are valued and encouraged.

Author Name : Giri Balasubramanian
Company : CARE2DATA
Address : 4th Floor, 40, II Main Road, R.A. Puram,
City / Postcode : Chennai – 600 028
Work Phone : +91-44-42606602
Email : giri@care2data.com
Web : WWW.CARE2DATA.COM